

The Relevance of Referring Expressions: the Case of Diary Drop in English

Katherine J. Scott

Thesis submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

University College London
June 2010

I, Katherine J. Scott confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.....

Abstract

This thesis offers a pragmatic analysis of subjectless sentences in non-null subject languages, focusing on English ‘diary drop’ (as in ‘Saw a good film yesterday’). In chapter 1, I survey the data and discuss existing syntactic analyses (Haegeman & Ihsane 1999, 2001). While these generally acknowledge the importance of pragmatic factors in an overall account, no detailed investigation of their contribution has been proposed.

In chapter 2, I consider subjectless sentences in child language, and suggest that relevance theory (Sperber & Wilson, 1986/95) can shed light on why such utterances occur. In chapter 3, I revisit the adult data, and having established that null subjects function as referring expressions, I consider two pragmatically-oriented approaches to the analysis of referring expressions: Accessibility Theory (Ariel, 1990) and the Givenness Hierarchy (Gundel, Hedberg and Zacharski, 1993). Both adopt the relevance-theoretic framework, but claim that relevance alone is insufficient to account for the data.

In chapters 4 and 5, I develop a relevance-based account of referring expressions, and argue that we can do without the machinery of Accessibility Theory and the Givenness Hierarchy on two assumptions: first, that referring expressions encode procedural as well as conceptual meaning (Blakemore 1987, 2002), and second, that this procedural meaning does not identify the intended referent by appeal to considerations of Accessibility or Givenness. An important implication of my account is that the choice of referring expression not only affects reference resolution but can also contribute to what is implicitly communicated by an utterance. I provide detailed evidence for this.

In chapter 6, I return to the original null subject data and show that my relevance-based approach sheds new light on how these utterances function in a non-null subject language. In Chapter 7, I draw general conclusions and revisit the conceptual-procedural distinction in light of the analyses proposed.

Table of Contents

Chapter 1: The Diary Null Subject in English	10
1.1 Introduction.....	10
1.2 Environments	11
1.3 Distribution and properties	13
1.3.1 Distribution.....	13
1.3.2 Syntactic properties.....	14
1.3.3 Associated characteristics.....	16
1.3.4 Null objects.....	17
1.3.5 Interrogative sentences.....	17
1.3.6 The question of embedded null subjects	18
1.3.7 The optionality of the null subjects.....	18
1.4 Diary drop in other languages	19
1.5 Other null subjects.....	20
1.5.1 Informal style deletions	20
1.5.2 Empty categories: <i>pro</i> and topic-drop analyses.....	22
1.5.2.1 Government and Binding empty categories.....	22
1.5.2.2 Anaphoric and pronominal: PRO	23
1.5.2.3 Anaphoric and non-pronominal: NP-trace.....	23
1.5.2.4 Pronominal and non-anaphoric: <i>pro</i>	24
1.5.2.5 Non-anaphoric and non-pronominal: Topic Drop.....	27
1.5.3 Child nulls as an analogue.....	30
1.6 Subsequent analyses	30
1.6.1 An antecedentless empty category.....	30
1.6.2 Pronoun ellipsis.....	35
1.6.3 Zero spell-out and a move towards pragmatics	36
1.7 Concluding remarks and the next steps	40
Chapter 2: Null Subjects in Child English.....	42
2.1 Introduction.....	42
2.1.1 The null subject phase	42
2.1.2 Competence or performance?	44
2.2 Competence Accounts	45
2.2.1 Advantages of a competence approach to child nulls	45

2.2.2	The <i>pro</i> hypothesis.....	46
2.2.3	The VP hypothesis	50
2.2.4	Topic drop.....	51
2.2.5	Truncation.....	53
2.2.6	Summary.....	54
2.3	Performance Accounts.....	55
2.3.1	General advantages of performance-based accounts	55
2.3.2	Processing accounts	57
2.3.2.1	Sentential complexity: a general processing approach.....	57
2.3.2.2	VP length and rightward complexity	58
2.3.2.3	A metrical approach.....	60
2.3.3	Testing the predictions	61
2.3.4	Comparing approaches	62
2.4	Discourse pragmatic approaches: two types of informativeness	63
2.4.1	The role of pragmatics.....	63
2.4.2	Informativeness 1: Greenfield & Smith (1976)	63
2.4.3	Informativeness 2: Allen (2000).....	65
2.4.3.1	Overview	65
2.4.3.2	The features	66
2.4.3.3	Informativeness and English null subjects.....	67
2.4.3.4	Problems and limitations of the informativeness account	68
2.4.4	A move towards relevance	71
2.5	Relevance Theory.....	72
2.5.1	An overview of relevance theory.....	72
2.5.2	Relevance and child nulls.....	76
2.5.3	Relevance instead of Informativeness.....	79
2.5.4	Concluding remarks	81
Chapter 3:	Referring and Accessibility.....	83
3.1	Introduction.....	83
3.2	Contextual accounts.....	83
3.3	Accessibility theory	85
3.3.1	Overview of Accessibility theory	85
3.3.2	A scale of Accessibility	86

3.3.3	Accessibility theory and relevance theory.....	89
3.3.4	Problems and limitations	91
3.3.5	Special uses and special problems	92
3.3.5.1	Special uses	92
3.3.5.2	Special problems.....	93
3.4	Gundel and colleagues: The Givenness hierarchy	94
3.4.1	Overview	94
3.4.2	The Givenness hierarchy and Grice	97
3.4.3	Limitations of a Gricean analysis	100
3.4.4	The Givenness hierarchy and relevance theory	103
3.5	Against a scalar analysis: the case of names.....	105
3.6	Concluding remarks.....	109
Chapter 4: Referring Expressions and Relevance Theory		111
4.1	Introduction.....	111
4.1.1	Arguments against a relevance-only account	111
4.1.2	The relevance theoretic comprehension procedure.....	115
4.1.3	The conceptual-procedural distinction and the role of procedural meaning.....	116
4.2	Explicitly communicated meaning.....	121
4.2.1	Determining the proposition expressed.....	121
4.2.2	Conceptual information and explicit meaning.....	123
4.2.3	Procedural information and explicit meaning.....	125
4.2.4	Applying the constraints.....	127
4.3	Beyond reference.....	131
4.3.1	Conceptual information and implicit meaning	131
4.3.2	Procedural information and implicit meaning	136
4.4	Concluding remarks.....	139
Chapter 5: Motivating the distinctions: demonstratives and stressed pronouns		140
5.1	Introduction.....	140
5.2	Demonstrative determiners	141
5.2.1	Existing analyses.....	141
5.2.1.1	Gestural / spatio-temporal uses	142
5.2.1.2	Anaphoric / discourse uses.....	143
5.2.1.3	Symbolic and emotional uses.....	145
5.2.1.4	Towards a unified approach.....	148

5.2.1.5 Kaplan: Pure indexicals, true demonstratives and intentions.....	151
5.2.2 A procedural analysis of demonstratives.....	153
5.2.3 Demonstrative determiners and explicatures.....	158
5.2.3.1 ‘This’ and ‘that’.....	158
5.2.3.2 Demonstrative uses.....	160
5.2.3.3 Indexical uses.....	161
5.2.4 Demonstrative Determiners and Implicatures.....	163
5.2.4.1 Explicit or implicit: a test.....	163
5.2.4.2 Demonstratives and inferential effects.....	165
5.2.5 Revisiting the demonstrative categories.....	167
5.2.6 ‘Special uses’ revisited.....	174
5.3 Stressed and unstressed pronouns.....	176
5.3.1 Relevance and accessibility.....	176
5.3.2 Factors affecting accessibility of interpretations.....	176
5.3.2.1 Order of mention.....	177
5.3.2.2 Syntactic position.....	177
5.3.2.3 Semantics of the main verb.....	177
5.3.2.4 Parallel function.....	178
5.3.2.5 Accessibility of contextual assumptions.....	179
5.3.2.6 Referring expressions and accessibility of interpretations.....	180
5.3.3 Stressed and unstressed pronouns: a relevance-theoretic account.....	181
5.4 Conclusion.....	185
Chapter 6: Null Subjects Revisited.....	186
6.1 Introduction.....	186
6.2 Null subjects as referring expressions.....	186
6.2.1 Accessibility and null subjects.....	186
6.2.2 Accessibility and avoiding pronouns.....	188
6.2.3 The Givenness hierarchy and null subjects.....	190
6.3 Relevance theory and overt pronouns.....	192
6.3.1 Overt pronouns as referring expressions.....	192
6.3.2 Types of pronouns.....	193
6.3.3 Pronouns: procedural and truth-conditional.....	194

6.3.4	A pronominal procedure.....	195
6.3.5	Third person pronouns	197
6.3.5.1	Gender and procedures	197
6.3.5.2	The role of contextual assumptions	205
6.3.5.3	Gender-redundant uses.....	208
6.3.5.4	Informative uses	209
6.3.6	The first person singular pronoun	210
6.4	Relevance theory and null subjects	214
6.4.1	Greater or different effects	214
6.4.2	Greater effects and different effects.....	215
6.4.2.1	The informal null	215
6.4.2.2	The pressurized null.....	218
6.4.2.3	The ostensibly vague null.....	220
6.4.3	Null subjects and poetic effects	222
6.5	Concluding remarks.....	224
Chapter 7: Conclusion.....		226
7.1	Introduction.....	226
7.2	Non-subject omissions.....	226
7.2.1	Determiners.....	226
7.2.2	Auxiliaries and the copula	230
7.2.3	Objects.....	231
7.3	Further null-subject data	234
7.3.1	Embedded and root nulls	234
7.3.2	Child nulls again	235
7.3.3	Imperatives	237
7.4	Concluding remarks.....	241
7.4.1	Revisiting procedural meaning	241
7.4.1.1	The vulnerability of procedural meaning to omission.....	241
7.4.1.2	The nature of procedural meaning.....	243
7.4.2	In support of a relevance-based account of referring expressions.....	245
7.4.3	Final remarks	247
References		248

Acknowledgements

First and foremost I would like to thank my supervisor Deirdre Wilson for being such an inspiration throughout my studies. Her patience and enthusiasm have been unstinting, and I could not have made it to this point without her help and support. Thanks also go to Robyn Carston for her help, advice and kindness throughout my time at UCL.

I would also like to thank Billy Clark for all the useful discussions, and for making teaching both interesting and fun. I'm hugely indebted to Catherine Wearing for support, advice, friendship and, of course, belaying.

I want to thank Alison Hall, Thiago Galery, Ingrid Lossius Falkum, Matthew Reeve, Jonny McIntosh, Akis Kechagias, Beata Zacharska and all the other UCL linguistics research students for the useful discussions, and for making the PhD room such a great place in which to work. I'd like to thank all the UCL linguistics academic and support staff, in particular Molly Bennett, for always being there to offer practical help and advice, and Hans van de Koot for being such an efficient and approachable head of department. Special thanks go to the members of the relevance reading group for all the useful and enlightening discussions.

I am indebted to the Arts and Humanities Research Council for the doctoral award which supported this research.

I'm hugely grateful to my good friends Clare, Sarah and Ellen and to all the softballers, particularly the Wolves, for keeping me sane and making me smile.

I want to thank my father for his practical support and for always believing that I had it in me.

Finally, I want to thank my mother, Marian. Although she died shortly before the completion of this thesis, her love and support have underpinned everything in my life, and she continues to be my inspiration.

All mistakes in this thesis are, of course, my own.

Chapter 1: The Diary Null Subject in English

1.1 Introduction

English is generally considered to be a non-null subject language. Whereas subject pronouns may be omitted in some languages (including Spanish, Italian and Japanese), they must be overtly realised in others (including English, Danish and French) (Chomsky, 1981; Jaeggli & Safir, 1989). So, whilst (1) is perfectly grammatical in Spanish, the corresponding subjectless sentence in English, (2a), is ungrammatical. The pronominal subject must be articulated, as in (2b).

(1) baila bien.

(2) a. *dances well.

b. He dances well.

This thesis examines a range of data that seem to contradict this generalisation, and considers some of the pragmatic factors affecting their interpretation and use.

In certain registers of English, speakers systematically produce utterances with missing or null subjects. Consider a sentence such as (3), written as the opening line of a diary entry.

(3) Got up late this morning.

It is not unusual for native speakers to produce such subjectless sentences when communicating in certain registers. In this chapter, I take a closer look at this phenomenon, examining some of the environments in which it occurs and the constraints on its use. Given the importance for syntactic theory of null-subject sentences in a non-null subject language, it is perhaps not surprising that most existing analyses have been heavily syntax oriented, and I will briefly outline some of the syntactic treatments proposed. However, the interpretation and acceptability of null-subject utterances varies considerably from situation to situation, and most syntactic theorists also appeal more or less explicitly to pragmatic factors to explain this

variation. My main aim in this thesis is to provide a pragmatic analysis of null subject utterances in English, focusing in particular on the case of diary drop.

Much research into the properties and distribution of null subjects in non-null subject languages has been undertaken by Liliane Haegeman and her colleagues (Haegeman, 1990a; 1990b; 1997; 2000; Haegeman & Ihsane, 1999; 2001). In several articles she has surveyed a wide range of examples and offered various syntactic analyses of the phenomenon. My discussion of the grammatical aspects of English null subject data is greatly indebted to her invaluable insights and analyses.

1.2 Environments

The term ‘diary drop’ has often been applied to the phenomenon of null subjects in written English. However, their distribution is not limited to diaries. We find examples of dropped subjects in a wide range of written material including emails, text messages, telegrams, postcards, note-taking and message boards¹. The diaries themselves can be fictional or non-fictional and examples can be found as far back as the diary of Samuel Pepys, as well as in contemporary works such as *Bridget Jones’ Diary*.

(4) Walked to Westminster hall. (Pepys, 1985, p. 610)

(5) Sat there for $\frac{3}{4}$ hour. (Woolf, 1985, p. 334)

(6) Went to the chemist to discreetly buy pregnancy test. (Fielding, 1996, p. 117)

Postcards, as in (7), telegrams, as in (8), and other written sources where space or time is restricted also provide a range of examples.

(7) Visited the castle yesterday. Wish you were here.

(8) A: Please attend first night my new play. Will hold two tickets for you. Bring friend if you have one.

¹ Many of the examples discussed in this thesis come from written sources. However, I will refer throughout to ‘utterances’ and to the ‘hearer’ and ‘speaker’. Unless specified, I do mean to imply any interesting distinction between written and spoken uses of language.

(9) B: Impossible attend first night. Will attend second night if you have one.²

(10) Contains nuts.

Certain first person narrative literary texts, whilst not explicitly diaries, employ the device of dropped subjects, as in (11)-(12).

(11) Went to the Argyll Lounge; good view of the harbour from there. Drank pints.
(Banks, 1996, p. 313)

(12) Saw a glowing red line in front of me, like a vein of burning blood, like lava, in front of me. Noise terrific. Smell of sulphur, something of that nature; smell of the devil, though I think that was just coincidence. Fell down. Half blind. Thought a bomb had gone off. Heard ringing, like the church bells all going on at once. Realised it was lightning. (Banks, 1996, p. 316)

The recent increase in electronic and online forms of written communication brings with it examples of diary-style nulls from emails, online message forums and text messages.

(13) Still have horrible cold. (personal email)

(14) Just ate a huge prawn ciabatta for lunch. (personal email)

(15) Was delivered promptly. (internet auction site message board)

(16) Very happy with dealer. (internet auction site message board)

(17) Am delighted with dress. (internet auction site message board)

(18) Will act on stage in 1st play. (Message on Twitter.com – K.Spacey 4th June 2009)

(19) Must have missed one another. (personal SMS/text message)

² Attributed to George Bernard Shaw and Winston Churchill respectively.

It has been observed (Massam & Roberge, 1989; Culy, 1996; Wharton, forthcoming) that in recipes and other instructional registers objects often remain implicit, as in (20) and (21).

(20) Roll out pastry and cut.

(21) Peel carrots and boil.

1.3 Distribution and properties

1.3.1 Distribution

The examples in (4)-(19) are drawn from a variety of sources. However, there are certain distributional features which they all appear to share. Whilst null subjects are most commonly understood as first person singular, they are not restricted to this interpretation. Examples (22) and (23) illustrate null subjects in the third person singular and first person plural, respectively, and Haegeman and Ihsane (1999, p. 132) give (24) as an example of a second person null subject.³

(22) Probably wants you to reply to his message. (personal email)

(23) Very excited about entering our 2nd phase of 10 year plan. (Message on Twitter.com K. Spacey 4th June 2009)

(24) Trouble is, Rebecca's stings are aimed so subtly at one's Achilles' heels, like Gulf War missiles going 'Fzzz Whoossh' through Baghdad hotel corridors, that never see them coming. (Fielding, 1996, p. 146)

Haegeman and Ihsane (1999) suggest that the relative rarity of second-person examples could be 'due to the fact that diary writing rarely addresses an interlocutor directly' (1999, p. 121). Examples of third person null subjects can be found with both referential (25) and expletive (26), and both human (27) and non-human interpretations.

(25) Was my mother. (Fielding, 2004, p. 5)

³ Perhaps this example is better analysed as involving the generic third-person singular pronoun 'one'.

(26) Is pissing it down. (Fielding, 2004, p. 10)

(27) Still has not woken up. (Fielding, 2004, p. 4)

(28) Will be so lovely having a boyfriend when it is warm. (Fielding, 1996, p. 137)

Consider the utterance in (29), taken from an online discussion board.

(29) Empathise about the masters thing. My dissertation is due in tomorrow. Can't wait!!! Nearly wasn't going in at all. Got burgled last night. Took a bag with my 3 copies of my dissertation in! Luckily dumped them outside – it's obviously not a very exciting read!!

Here the speaker drops all but one of her pronominal subjects. Alongside the first person singular (and possibly plural), we also have examples of the speaker dropping third person pronominals (when referring to the burglars).

1.3.2 Syntactic properties

Haegeman (1990a) presents clear evidence that the null subjects in diary texts are represented syntactically. Consider the examples in (30) and (31)

(30) (ec)_i am going to try and behave myself_i. (personal email)

(31) (ec)_i walked there PRO_i feeling light and airy. (Smart, 1991, p. 15)

In (30), the empty category in subject position binds the reflexive pronoun: it agrees with 'myself' in number, gender and person. In example (31), taken from Haegeman (1997), the null subject controls PRO in a non-finite clause. Both examples demonstrate that the null subject is syntactically active.

This syntactic behaviour contrasts with that of the non-overt argument in passive constructions. Such arguments are understood, but are not syntactically active.

(32) They put the book on sale.

(33) They put the book on sale and it was sold last week.

(34) They sold the book to themselves.

(35) *The book was sold to themselves.

In the passive clause in (33), the agent is not overtly expressed but is understood. However, (35) demonstrates that, in contrast to the diary null subject in (30), the understood passive subject is unable to bind a reflexive. So whilst it is possible to have empty argument slots for which a referent is understood without being syntactically represented, this does not seem to be the situation with the diary nulls.

There is also evidence that the diary nulls are represented at the semantic level. Consider (36).

(36) (ec) saw John at the weekend.

This sentence and others like it are interpreted as if they have a subject: indeed, throughout the texts containing diary style nulls, we see overt forms alternating with the null subjects. From a semantic perspective, the theta criterion (Chomsky, 1981; Haegeman, 1994, p. 54) stipulates that there must be a one-to-one mapping between theta roles and arguments. A transitive verb such as 'see' assigns two theta roles: a perceiver and a perceived. We would therefore expect a sentence in which it occurs to have two arguments. However, (36) is perfectly acceptable in a diary context. In this example, the theta role associated with the role of the entity perceived is assigned to the overt argument 'John', but the sentence is also understood as involving someone or something doing the perceiving. In order to avoid a violation of the Theta Criterion, it is therefore necessary to hypothesize that there is some sort of empty category (ec) in the subject position, to which the perceiver theta role is assigned.

The hearer's goal in interpreting an utterance is to infer the speaker's meaning from the linguistic properties of the sentence uttered, together with background information. To achieve this goal, reference must be assigned to the missing argument, marked here as (ec). The logical form of (37), which contains a null subject, should therefore be identical to that of a parallel sentence containing an overt pronoun, such as (38).

- (37) a. ec saw a good film.
b. LF: X saw a good film.
- (38) a. He saw a good film.
b. LF: X saw a good film.

Semantically, then, the null subjects function as non-overt pronouns.

The evidence surveyed above suggests that the null subjects in diary-style texts are syntactically and semantically active non-overt NPs. They can carry theta roles, and they enter into syntactic relations in much the same way as overt NPs, but have no realization at the level of phonological form (PF). Thus, they appear to be functioning as some kind of empty grammatical category. Taking this as a starting point for our analysis, we are led to ask what the nature of this empty category might be. In the next section I attempt to answer this question, comparing the behaviour and distribution of the diary-style nulls with those of the empty categories that are independently motivated in the Government and Binding grammatical framework (Chomsky, 1981; Haegeman, 1994).

1.3.3 Associated characteristics

The texts and registers in which we find null subjects commonly also display other deletions. The example from a telegram in (8) contained instances of unarticulated determiners, prepositions and even the infinitival 'to', and we see similar patterns in the diary texts. Determiners, both definite, as in (39) and (41), and indefinite, as in (40), are frequently left non-overt in such texts, and in (40) we also have an instance of a non-overt possessive pronoun.

- (39) That would indeed be sick in manner of Oedipus. (Fielding, 2004, p. 5)
- (40) Also managed to conceal coat by rolling it into ball to create pleasing sense of having been in for hours. (Fielding, 2004, p. 8)
- (41) Dialling tone seems normal. (Fielding, 2004, p. 45)

Auxiliary verbs and the copula also frequently remain non-overt straight after a null subject, as in (42)-(43).

(42) Maybe could read in car when at traffic lights. (Fielding, 2004, p. 13)

(43) If not working might mean everything is fine. (Fielding, 2004, p. 45)

1.3.4 Null objects

Although it is most frequently subject arguments which remain unarticulated, there are examples of null objects in many of the diary-style texts.

(44) Eventually locate under clothes from wardrobe. (Fielding, 1996, p. 93)

(45) Search for hairbrush. Locate in handbag. (Fielding, 1996, p. 92)⁴

(46) I have xmas cards in the office. If you wish to send one to a supplier please come to the office and collect. (personal communication)

Whilst such examples exist, there is a clear asymmetry in the distribution of null subjects as compared with null objects. A comprehensive analysis of the data should be able to account for the fact that both null subjects and null objects are attested, while also addressing the striking asymmetry in their distribution.

1.3.5 Interrogative sentences

It has been observed (Haegeman & Gueron, 1999; Haegeman & Ihsane, 1999; 2001) that there appear to be syntactic constraints on the distribution of non-overt subjects in the diary-style registers. According to Haegeman and Gueron, null subjects may only occur in declarative sentences, since they are incompatible with both *wh*-movement and subject-auxiliary inversion. They give the following as examples of attested subjectless sentences, which are judged to be ungrammatical when interpreted as interrogatives.

(47) *Ought (I) to resign?

⁴ Examples (44), (45), (50) and (51) are taken from Haegeman & Ihsane (1999)

(48) *Am (I) jealous?

(49) *What can (I) say?

In a survey of diary-style texts, Haegeman and Ihsane (1999) find no examples of null subjects occurring in *wh*-preposing or subject-auxiliary inversion environments. However, they acknowledge that given the relative rarity of interrogative sentences in these texts, 'the absence of the null subject may be a sampling accident' (1999, p. 134).

1.3.6 The question of embedded null subjects

Rizzi (1994), Haegeman (1997), and Haegeman and Gueron (1999) assume that null subjects are restricted to root clauses. In section 1.6.1 I will outline an analysis based on this assumption which also claims to account for the lack of null subjects in interrogative sentences. However, Haegeman and Ihsane (1999; 2001) discuss data which casts doubt on the validity of this assumption. Certain recent diary-style texts contain null subjects in embedded positions:

(50) But even that is inadvisable since am fat. (Fielding, 1996, p. 17)

(51) Deceiving her impossible as would be unsuccessful in long run. (Ginsburg, 1995, pp. 49-50)

(52) Wonder if should quickly ring Mark Darcy to tell him where am going?
(Fielding 2004, p. 13)

Whilst some texts contain only root null subjects (RNS), all those that include embedded null subjects (ENS) also display root nulls along with the other omissions. Setting aside theoretical considerations, it feels intuitively as if they are part of the same phenomenon and ideally I hope to settle on an analysis which covers both sets of data.

1.3.7 The optionality of the null subjects

In texts and registers which feature the diary-style null subjects, it is notable that the omission of the articulated form is always optional. There appear to be no cases where it would be ungrammatical to use an overt form, and the overt and non-overt subjects

alternate, apparently freely. Thus, an utterance such as (53) can be followed by (54) without either being judged ungrammatical in the diary-style register.

(53) Seems rude not to reply. (Fielding 1996, p. 44)

(54) I'll just send him a tiny friendly message. (Fielding 1996, p. 44)

However, despite their grammaticality and apparent synonymy, there are subtle stylistic differences between these two utterances, which an adequate pragmatic analysis should help to explain.

1.4 Diary drop in other languages

So called diary-drop is also attested in other non-*pro*-drop languages. Haegeman (1990b; 1997) and Haegeman and Ihsane (2001) discuss some examples from French, such as (55), whilst Haegeman (1990b, pp. 171-172) provides examples from Dutch, as in (56), German, as in (57) and Danish, as in (58).

(55) Suis tellement énervée que me suis assise sur la télécommande.

Am so nervous that me am seated on the remote control.

' am so nervous that I have just sat down on the remote control.'

(Stroumza, 1998, p. 182)

(56) Vind, in Gryll Grange, ook een erg aardige uitdrukking die.

Find, in Grylle Grange, another nice expression which.

(57) Beschloss, ihn zu kaufen, änderte diesen Beschluss...

Decided it to buy, changed this decision...

(58) Har blot ingen tid at scrive i.

Have just no time to write in.

A detailed contrastive analysis - which lies beyond the scope of the present thesis - might shed valuable further light on the phenomenon of diary drop.

1.5 Other null subjects

Although English is traditionally classified as a non-*pro*-drop language, there are several situations in which subjects may remain unarticulated. In this section I examine these data and consider whether they pattern with the diary-style nulls described above. I will also consider whether diary drop is best approached in terms of a *pro*-drop or topic drop analysis as I compare the characteristics of the diary nulls with independently motivated empty categories.

1.5.1 Informal style deletions

With many utterances of spoken English, some expected initial material is left unpronounced. This is especially common in isolated utterances and in fast, casual speech. Consider the examples in (59)-(67), where the missing material is shown in brackets.

(59) (I will...) See you soon.

(60) (I...) Can't find my earrings.

(61) (Have you...) Seen the paper today?

(62) (Is the...) Post here yet?

(63) (Are you...) Coming for a drink?

(64) (A...) Fine friend you turned out to be! (Napoli, 1982).

(65) (The...) Cat's been sick everywhere.

(66) (Of...) Course you can!

(67) (I expect...) 'spect so.

Data such as these are discussed by Napoli (1982) and Zwicky & Pullum (1983). The roles played by sentence position, constituency and stress in deletions of this kind are considered, and both analyses conclude that the phenomenon cannot be syntactic in nature. For Napoli, informal speech may involve a phonological rule that 'deletes unstressed (or lightly stressed) initial material' (1982, p. 99). Zwicky and Pullum, on the other hand, conclude that the informal style deletions are the result of a morphophonemic process.

How far might these analyses be useful to us in our investigation of diary-style null subjects? Whilst there are certain similarities between the data in (59)-(67) and the diary null subjects, there are also notable differences. As Napoli points out, the informal-style spoken deletions, like the written examples, are syntactically represented. They trigger agreement on the verb, co-refer with the subject of a tag question, as in (68), and can control the subject of an infinitival clause.

(68) Can't sing a note, can he?

However, the restriction of this type of deletion to initial position is far stricter than in the diary-style cases, and the associated deletion of determiners and copulas is also restricted to first position. As Haegeman and Ihsane (2001) note, an utterance such as (69), from Thrasher (1977), whilst acceptable in the written diary-style texts, is ungrammatical in all registers of spoken English, where the determiner in the object noun phrase must be overtly articulated.

(69) Damn dogs are taking over city.

Despite the differences, in both the diary-style nulls and the casual spoken deletions, the hearer must recover the content of the unpronounced material in order to understand the speaker's meaning. If, as I will argue, the vulnerability of a given item to deletion depends partly on the nature of the encoded material, then we would expect to see some overlap in the pragmatic factors affecting deletion in the two different registers.

1.5.2 Empty categories: *pro* and topic-drop analyses

1.5.2.1 Government and Binding empty categories

Given the above evidence that the diary null subjects are syntactically active, a natural move in investigating them further is to compare them with other empty categories. Within the Government and Binding framework (Chomsky, 1981; 1982; Riemsdijk & Williams, 1986; Haegeman, 1994; Black, 1996), two binary feature specifications (+/- pronominal and +/- anaphoric) combine to yield four independently motivated empty categories. An anaphoric element is referentially dependent on an antecedent which provides the necessary information for its interpretation. Pronominal elements, on the other hand, substitute for and function as nouns or noun-phrases.

Table 1: GB Empty Categories

Anaphoric	Pronominal	Empty Category
+	+	PRO
+	-	NP-trace
-	+	<i>pro</i>
-	-	WH-trace

Within the Government and Binding framework, all empty categories are subject to the Empty Category Principle (ECP), given here in (70).

- (70)
- a. Formal licensing: An empty category must be governed by an appropriate head.
 - b. Identification: An empty category must be chain-connected to an antecedent.

For an empty category to be licit, it must be both licensed and identifiable, so that both clauses of the ECP are satisfied. In the remainder of this section, I consider the suitability of the different empty categories in the table above for analysing diary-style nulls. Several accounts of the non-overt English subjects align the nulls with one or other of these categories, and I will now consider the advantages and disadvantages of such accounts in more detail (Haegeman, 1990a; Bromberg & Wexler, 1995).

1.5.2.2 Anaphoric and pronominal: PRO

PRO ('big pro') is the non-overt subject of non-finite clauses. Within the GB framework, the Extended Projection Principle (EPP) states that every sentence must have a subject, and the Case Filter states that only case-marked NPs may be phonetically realized. In the case of non-finite clauses, these two principles combine, leaving us with a syntactically active yet phonetically null subject: PRO.

(71) PRO to pass one's exams is something to be proud of.

According to the EPP, since (71) is a sentence, it must have a subject. However, the verb in its non-finite form is unable to assign case to the subject position. The subject position is therefore not case-marked and so the subject, although obligatory, cannot be phonetically realized. Hence, PRO only occurs in non-finite clauses and does not alternate with overt NPs. An overt form in the PRO position would lead to a case filter violation, as in (72).

(72) *John to pass his exams is something to be proud of.

The overt subject, 'John', is unable to receive case from the infinitival verb, and so cannot be overtly realised. The diary null subject, on the other hand, occurs in finite clauses and, as noted above, is found alternating with overt arguments. The first of the independently motivated empty categories, PRO, is therefore ruled out as a possible analysis for the diary null subject.

1.5.2.3 Anaphoric and non-pronominal: NP-trace

NP-traces have the features [+anaphoric] and [-pronominal], and are found when movement takes place to an argument position, for example, in cases of passivization and raising, as in (74).

(73) There seems to be a man at the door.

(74) A man_i seems t_i to be t_i at the door.

As was the case with PRO, NP-traces are only found in non-finite clauses, and do not alternate with overt subjects. They can therefore also be ruled out as a possible analysis of the diary null subject.

1.5.2.4 Pronominal and non-anaphoric: *pro*

In traditional *pro*-drop languages, such as Italian, pronominal subjects may remain non-overt. In their place we find *pro* - ‘little pro’, the empty category with the features [+pronominal] and [-anaphoric]. Like the diary null subject, *pro* represents a syntactically active understood subject which may alternate with an overt form in certain circumstances. It has therefore been suggested that some speakers of English have a ‘*pro*-drop dialect’ that is available to them in certain registers of written and spoken language. An alternative suggestion might be that English is undergoing a change from being a non-*pro*-drop to a *pro*-drop language. However, although the data might, on the surface, seem to point to such explanations, on closer examination the diary null subject begins to look less and less like an instantiation of *pro*-drop.

The availability of *pro* as an alternative to an overt subject pronoun is generally treated as varying parametrically across languages. As we saw in example (1) from Spanish, if the parameter is positively set, subjectless sentences are grammatical, while the parallel sentence in a non-*pro*-drop language such as English is ungrammatical, as in (2a). Languages in which the *pro*-drop parameter is positively set generally display a bundle of associated properties. If the diary null subjects were instances of *pro*, with some sort of parametric change taking place in English, then we would expect to see the associated properties also occurring in the diary-style texts. In *pro*-drop languages, a definite subject may occupy a post-verbal position, as in the Italian example (75) taken from Haegeman (2000, p. 136).

(75) Ha telefonato il deano.

has telephoned the dean.

‘The dean has telephoned.’

(76) *Has telephoned the dean.

The corresponding sentence in English, (76), is not grammatical in either the core grammar or the diary drop registers on an interpretation where ‘the dean’ is the subject. Similarly, consider (77) and (78), again taken from Haegeman (2000).

(77) *Who_i do you think that t_i has come?

(78) Chi_i credi che abbia telefonato t_i?

Who believe you that have (subj) telephoned?

‘Who do you think has called?’

Sentence (77) is unacceptable in all registers of English due to a *that*-trace violation. However, in traditional *pro*-drop languages, an embedded subject may move across an overt complementizer, as illustrated by the grammatical Italian sentence in (78). Moreover, in *pro*-drop languages expletive subjects are always non-overt. In Italian a construction with an overt expletive, as in (79), from Haegeman (1997, p. 236), is ungrammatical.

(79) *Ciò piove.

It is raining.

However, throughout the registers that display diary-style null subjects, we find overt expletives alternating with non-overt expletives, as in (80) and (81) below.

(80) Is relief to have fag in open and not to be on best behaviour. (Fielding 2004, p.4)

(81) It was great. (Fielding 2004, p. 33)

Finally, and perhaps most dramatically, Italian style *pro*-drop is traditionally associated with a rich inflectional paradigm. It is often claimed that the rich inflection in Italian both licenses and identifies the empty category (Rizzi, 1986; 1994; Haegeman & Gueron, 1999, pp. 399-400), satisfying both clauses of the ECP (70). The English

inflectional system remains consistently poor across the varying registers, raising the question of how the empty subject is identified (Adger & Harbour, 2008, pp. 5-7).

In sum, if the occurrence of null subjects in the diary registers signalled a switch in the *pro*-drop parameter, then we would expect these associated properties to also be present. This is not the case. However, the absence of these characteristics is not the only evidence pointing us away from a *pro*-drop analysis. Not only do the diary-style texts lack the properties generally associated with *pro*-drop languages, but the characteristics which commonly co-occur with the diary nulls (see 1.3.3) are not generally present in *pro*-drop languages. Similarly, the apparent constraints on the distribution of the diary nulls, (see section 1.3) do not apply to the null subjects in *pro*-drop languages. Thus, Italian style *pro*-drop occurs in interrogative sentences, both root and embedded, and is freely compatible with *wh*-preposing, as in (82), and subject-auxiliary inversion, as in (83). It is also found with argument and predicate topicalization, as in (84)-(85).

(82) Quando [*pro*] tornerà?

When return-future-3sg.

'When will he/she return?' (Haegeman, 2000, p. 140)

(83) Tornerà [*pro*] presto?

Return-future-3sg soon?

'Will he/she return soon?' (ibid)

(84) Questo libro, non lo voglio.

This book *non* it want-I.

'This book, I don't want.' (Haegeman and Guéron, 1999, p. 621)

(85) Intelligente, non è.

Intelligent, non be-3sg.

'intelligent, he isn't.' (ibid)

Along with the distributional differences and the absence of the associated features that we would expect to see with a *pro*-drop dialect, there are also conceptual problems with a *pro* analysis of the diary style nulls.

If the diary null subjects were instances of *pro*, then some kind of register-specific parameter resetting would have to be taking place. The speaker would then have two grammars: a core grammar in which null subjects are ungrammatical and a peripheral one in which they are allowed. Haegeman (1990a) considers this possibility and suggests that such a register-driven resetting may be possible given enough exposure to relevant evidence, and once the core grammar has been firmly established. In a later article, however, (Haegeman, 2000) she identifies a range of problems with this account. The argument for treating the diary null subjects as instances of *pro* would be greatly strengthened if there were evidence of register driven parameter resetting elsewhere in the grammar. However, no such evidence seems to exist. We do not see, for example, languages with VO core word order changing to OV in a particular register. To justify a parameter-resetting approach, we would therefore have to show what makes the *pro*-drop parameter special in this respect, and why no other parameters vary in this way.

This combination of otherwise unattested parameter resetting and the absence of the expected associated features severely weakens the case for analyzing diary null subjects as instances of *pro*. So whilst at a first glance the diary nulls seem to pattern with the *pro* of null subject languages, an analysis on these lines would bring into question much that is commonly assumed of the classic *pro*-drop languages. I will revisit the possibility of a *pro*-drop analysis and what we can learn from it in the next chapter, when I look at the child null subject in English and its similarities with the diary style nulls.

1.5.2.5 Non-anaphoric and non-pronominal: Topic Drop

The remaining empty category in the GB typology carries the features [-anaphoric] and [-pronominal] and is the trace of A'-movement. This type of movement, for discourse

related purposes such as topicalization, focus or question formation, leaves a trace in its base position. Movement of this kind is widespread in English, including for the purposes of overt topic preposing, as in (86).

(86) CP[John_i IP[I spoke to t_i yesterday.]]

In languages such as Portuguese, this kind of movement may take place whilst the topicalised constituent remains non-overt.

(87) A Joana viu na televisao ontem a noite.

Joana saw on television last night.

‘Joana saw him/her/it on television last night.’

(88) CP[TOP_i IP[a Joana viu t_i na televisao ontem a noite.]] (Rizzi, 1986, p. 513)

The topic in [spec, CP] is an empty operator, which binds the object trace of ‘viu’. Identification of this empty operator then relies on some salient element in the discourse, in this case the person who is the current topic of the discourse.

It has therefore been proposed (Haegeman, 1990a; Bromberg & Wexler, 1995) that the diary drop empty category is the A' trace of a non-overt topic antecedent. Null operators are indeed attested in English under certain conditions (Haegeman & Ihsane, 1999):

(89) I need a friend [OP_i [I can rely on t_i]]

The object trace of the embedded verb ‘rely’ is bound by a non-overt antecedent, the null operator OP, and this in turn is identified by the adjacent head, ‘a friend’. On this account, the representation of a diary-style sentence such as (90) would be as in (91).

(90) Could stand it no longer.

(91) CP[OP_i[IP[t_i could stand it no longer.]]

As Haegeman (1990a, p. 176) sums up:

there is a trace of *wh*-movement (specifically topicalization...) in the subject position of the sentence...and this trace is coindexed with the moved non-overt or zero discourse TOPIC.

Bromberg and Wexler (1995) argue that this analysis works well to explain the incompatibility of the null subject with *wh*-preposing. They start from the assumption that both topic and *wh*-preposing target the same tree position and cannot, therefore, occur simultaneously. However, this argument breaks down once the underlying assumptions are examined. As Haegeman (2000) points out, several linguists (Reinhart, 1981; Rizzi, 1997), have argued for an articulated rather than unitary CP projection. This articulated CP is made up of various functional projections with specific roles and functions, including a singular Focus Phrase, which is the landing site for *wh*-movement, and a recursive Topic Phrase that, in turn, is the landing site for topicalized constituents. This independently motivated articulated CP therefore undermines one of Bromberg and Wexler's main supporting arguments for the topic drop analysis.

Two further substantial objections to this analysis have also been raised (Haegeman, 2000). Firstly, there is the matter of subject/object asymmetry. As discussed above, the texts containing cases of subject drop show few, if any, corresponding instances of object drop. By contrast, in languages, such as Portuguese, that display topic-drop, the null topic is often, if not typically, an object. To justify the topic-based analysis, this asymmetry in the diary drop data would need to be accounted for by some other means. Secondly, if the null subject in these registers were in fact a topic, then we would expect only referential, non-expletive subjects to be found in this position. However, non-referential null subjects are abundant in these texts:

(92) Must be an accident up ahead. (Schmerling, 1972, p. 582)

(93) Isn't much we can do about it. (Thrasher, 1977, p. 44)

(94) Seems weird to have been so close during the year...(Fielding, 1996, p. 296)

In conclusion, the analysis of the diary-style null subjects as instances of topic drop fails to be fully compatible with the distributional properties observed, and I will therefore not adopt it here.

1.5.3 Child nulls as an analogue

Having rejected the analyses based on the empty categories identified by Government and Binding theory, alternative possibilities for how we might account for the data must be considered. It is well documented (Hyams, 1986; 1992; Bloom L. , 1970; Bloom P. , 1990) that children acquiring a non-null subject language go through a stage in which they persistently produce subjectless sentences. It has been observed (Haegeman, 2000; Haegeman & Ihsane, 2001; Rizzi, 1994) that the diary-style null subject utterances share syntactic and distributional properties with the utterances produced by children during this stage of acquisition. For example, children in this stage frequently also omit other parts of speech, including articles, auxiliaries and copulas. Unlike the adult casual speech examples discussed in 1.5.1, these omissions are not restricted to initial position. Haegeman & Ihsane (2001, p. 331) give examples of non-initial article omission, as in (95), and auxiliary omission, as in (96).

(95) Paula play with ball. (Paula, 1, 9)

(96) Daddy gone. (Hayley, 1, 8, as cited in Radford (1996, p. 44))

In this respect, the diary style null subjects seem to pattern more with the child nulls than with the informal register nulls in adult English. In the next chapter, I take a closer look at the phenomenon of null subjects in the speech of children acquiring non-null subject languages. Meanwhile, the observed similarities between diary-style nulls and child omissions have been influential in the development of subsequent syntactic analyses of the diary null subject. I consider these accounts in the next section.

1.6 Subsequent analyses

1.6.1 An antecedentless empty category

As we have seen, according to Government and Binding theory, the four types of empty categories can be distinguished by breaking them down into the primitive features of [+/- anaphoric] and [+/-pronominal]. None of the types of empty category that emerge fits exactly with the distribution and behaviour of the null subject found in the diary style texts. However, building on the similarity between diary null subjects and child null subjects, and drawing on work by Lasnik & Stowell (1991), Rizzi (1994) proposes a further option. He argues that another feature should be added to distinguish sub-

categories within the category of A' traces. Ordinary operator constructions such as questions, as in (97), differ from null operator constructions, such as (98), in both their interaction with weak crossover effects and their interpretative properties.

(97) John wonders who_i to please t_i .

(98) John is easy OP_i to please t_i .

In (97), quantification of the trace of A'-movement (t) ranges over a 'possibly nonsingleton set' (Rizzi 1994, p. 158): there can be more than one person corresponding to the interrogative 'who'. In (98), however, the identification of the trace is fixed by that of the antecedent, John, and so cannot range over a non-singleton set. Rizzi concludes that only when a trace is bound by a genuine quantifier, as in (97), can it be said to be a variable. When the trace is bound by a non-quantificational empty operator, as in (98), Rizzi classifies it as a 'null constant'. Rizzi incorporates this difference into the taxonomy of empty categories by adding a [+/- variable] feature. A' traces resulting from *wh*-movement typically quantify over a 'nonsingleton set' and have their variable feature positively specified. In contrast, the traces resulting from the null operator movement never quantify over a nonsingleton set, and so are assigned the feature value [- variable]. Recall that, according to the ECP (see (70) above), all traces are subject to an identification requirement. As Haegeman and Ihsane (1999, p.124) put it:

The content of an empty subject is identified by a c-commanding antecedent, i.e. an antecedent which precedes the empty category and is hierarchically higher in the structure.

So, in order for this analysis to hold in these null subject cases where there is no antecedent, Rizzi has to modify the identification clause of the ECP. He does this as follows:

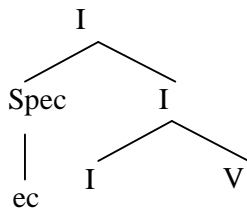
(99) A non-pronominal empty category must be chain connected to an antecedent *if it can be*.

It follows that an exception to the identification condition is allowed if the empty category is in the highest position in the structure. This analysis is based on the assumption that null subjects are restricted to root clauses. If the null subject is

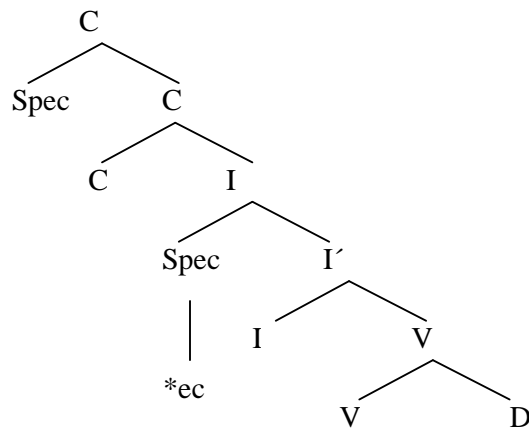
anywhere other than in the root clause, then there will be a potential site for the antecedent, and so one must be supplied and chain connected to the empty category. In the case of a root null subject, there is no potential antecedent position and it can therefore remain antecedentless.

Rizzi's work is particularly concerned with the null subjects found in child language, and suggests that whilst root clauses must be projected to CP in the adult grammar, they may be truncated in the child grammar. If, in the child's grammar, finite clauses need not be projected past IP, there will be no potential site for an antecedent to sit in these cases. According to (99), such empty categories may remain antecedentless, and so will have the structure in (100). Haegeman (2000) considers parallel cases in adult language and suggests that a similar process of truncation is taking place in abbreviated registers. In the core grammatical registers that do not allow truncation, the presence of a CP projection means that an antecedent position is available, as in (101), and root null subjects are therefore not grammatical.

(100)



(101)



What, then, might be driving the truncation of structure and subsequent licensing of the null subject forms in child versions of non-*pro*-drop languages and in the diary style registers of the same languages? The child null forms will be discussed in more detail in the following chapter. However, Haegeman (1997) draws on work from Rizzi (1994; 1997) to offer a possible solution based on the interaction of two principles:

(102) Root = CP.

(103) Avoid Structure.

Principle (102) is based on the assumption that C carries information about illocutionary force, and that since all clauses must have illocutionary force, they must all project to CP level. Rizzi (1994, pp.162-163) paraphrases this as a claim ‘that we normally speak through propositions, not fragments of propositions’ and that ‘the root category is the canonical structural realization of the proposition’. Principle (103) is based on economy considerations: it ensures that syntactic structure is not projected unless it is necessary, and cannot be posited purely to avoid rule violation (Rizzi, 1997, p. 314). These two principles will on occasion be in conflict, as Haegeman explains:

because of [(102)], a fundamental structure building principle, CP must be projected. By [(103)], on the other hand, syntactic structure cannot be added, for instance, to salvage a potential empty category principle violation. (Haegeman, 1997, p. 243)

In ‘normal’ adult English, rule (102) wins most of the time, and as a consequence each clause must project to the CP level. When a CP level is projected, there will automatically be a slot for an antecedent, and according to (99) it must be filled. However, Haegeman (1997; 2000) and Rizzi (1994) suggest that under certain circumstances, including both child language and diary style registers, considerations of economy become more important, and (102) may be overruled by (103):

In view of the fact that economy considerations (‘brevity’) clearly play a part in the abbreviated registers, one might then suggest that there too [(102)] is overruled by economy considerations [(103)]. (Haegeman, 1997, p. 246).

If the clause is only projected as far as the IP level, there is no higher position in which the antecedent can sit. The diary null subject is therefore ‘an antecedentless empty category in [Spec, IP], with CP being truncated’ (Haegeman 2000, p. 147), and it is claimed that this truncation is characteristic of abbreviated registers such as diaries, note taking and casual speech.

Rizzi's account predicts that null subjects will always be the leftmost constituent in the structure. Haegeman (1997; 2000) demonstrates that this is not always the case. She provides examples of adjuncts preceding the null subject, as in (104), and contrasts this with ungrammatical cases of argument preposing, illustrated in (105).

(104) At night sent a packet to London. (Pepys, 18 April)

(105) *More problems don't need. (Thrasher, 1977, p. 83)

To account for these data, Haegeman offers a reinterpretation of Rizzi's original analysis. Firstly, Haegeman (1997, p. 251) presents independent motivation for decomposing the CP into a ‘set of functional projections’, with truncation free to apply at any level. In the case of adjunct preposing, the empty category has moved to the specifier of an agreement projection and so actually precedes the adjunct, as in (106).

(106) ec_i at night t_i sent a packet to London.

Whilst movement can bypass adjuncts, it cannot bypass arguments, and so (105), with the structure in (107), is ungrammatical.

(107) * ec_i more problems t_i don't need.

In sum, the interaction of the rules in (102) and (103) could go some way towards explaining what drives the dropping of subjects in the given diary contexts. In child language, we may suggest that the relative importance of the two rules has not yet been fully acquired, while in the adult null subjects we may look to the issue of economy for motivation. Root null subjects tend to occur in casual, informal, abbreviated registers. Newspaper columns and fictional diary writing are often subject to word limits, and text messages and e-mails may also be governed by time and/or length restrictions. In the adult, non-core registers that display null subjects, the pressure to abbreviate may

therefore prioritize an economy-based principle such as (103) to the point where it takes precedence over the rule in (102).

This account satisfies a version of the ECP with the second clause modified as in (99), but how does it account for the other properties associated with the diary-style nulls? The analysis seems to work well to explain the absence of subject drop with *wh*-movement. For *wh*-preposing to take place, a CP level must be projected, providing a potential antecedent site which, according to (99), must be used. Similarly, this account rules out subject-auxiliary inversion constructions occurring with a null subject, since with no CP projection, there is no C position available as a landing site for the auxiliary after inversion. It therefore seems satisfactory as an explanation for the distribution of the root diary null subjects, whilst simultaneously treating these null subjects as a possible adult analogue of the early null subject in child language⁵.

Each of the analyses considered in this section was designed to account for, and in most cases relied upon, the assumption that null subjects are restricted to root position and so do not occur as the subjects of embedded clauses. Although Haegeman (1997; 2000) and Horsey (1998) address the issue of null subjects appearing with preposed adjuncts, the assumption has remained that the missing subject cannot appear in an embedded environment. As outlined in section 1.3.6 above, this assumption has been questioned in recent work, and this has led Haegeman & Ihsane (1999; 2001) to offer a still further analysis, in which the phenomenon is characterised as a form of pronoun ellipsis.

1.6.2 Pronoun ellipsis

Given the embedded data discussed by Haegeman and Ihsane (1999; 2001), the antecedentless empty category analysis (Haegeman, 1997; Rizzi, 1994; 1997; Horsey, 1998) appears to be inadequate, since it predicts that such data will be unacceptable. However, both the root and embedded null subjects are otherwise identical and interpreted in the same way, and it seems preferable from a theoretical point of view to provide a single account that covers both. Accordingly, Haegeman & Ihsane (2001) propose that the two dialects which differ as to the grammaticality of embedded null subjects are not the output of two different grammars with differing parameter settings

⁵ With the development of the minimalist framework (Chomsky, 1995), it became necessary to adjust this analysis in line with the new and different assumptions introduced by this framework. Horsey (1998) offers a reinterpretation of Haegeman's analysis in these terms. However, I will not go into the details of this proposal here, since it covers much the same data as Haegeman and as such does not provide further insight into the nature of the diary null subjects.

of the *pro*-drop (or other) parameter, but are rather the output of a single grammar with an optional pronoun ellipsis rule. Haegeman & Ihsane support this analysis by providing evidence that registers displaying the diary-style null subjects also often reduce reflexives, again ellipting the pronominal part, as in (108)-(110).

(108) trying to whip up frenzy of emotion in self about end of war. (Fielding 1996, p. 121)

(109) brace self for touchdown. (Emma D. May, p. 11)⁶

(110) Have cigarette to cheer self up. (Fielding 1996, p. 92)

As discussed above, empty categories are subject to the ECP: they must be licensed and their content must be identifiable. On their pronoun ellipsis account, Haegeman and Ihsane (1999) propose that diary-style null subjects are ‘licensed by a specifier-head relation with AGR’ (1999, p. 143). They also propose that in the relevant registers, the usual constraints on identification are relaxed: ‘identification may be achieved either by an antecedent in the matrix clause or by a discourse antecedent’ (1999, p. 142). As Haegeman and Ihsane point out, a pronoun ellipsis account of this sort predicts that objects too should be omitted. The examples in (44)-(46) show that object omission does occur in these registers. However, it is far less common than subject omission, and this asymmetry is not directly addressed by the pronoun ellipsis account.

1.6.3 Zero spell-out and a move towards pragmatics

As mentioned above, an association is often made between the richness of a language's inflectional paradigm and the setting of the *pro*-drop parameter. European languages such as Italian and Spanish on one side, and English and French on the other, typically exemplify the divide. However, there is also a class of languages in which null subjects occur but which lack the expected rich inflectional paradigm, Chinese and Japanese being much discussed examples (Huang, 1984; Jaeggli & Safir, 1989). Several analyses of this phenomenon have been offered. The fact that argument omission in these languages is more widespread than in the classic *pro*-drop languages, since it is possible for any pronominal, has led to it being termed ‘rampant’ or ‘radical’ *pro*-drop.

⁶ Taken from ‘Bad Trip, the diary of Emma D. May’, a 1000 word fictional diary in *The Independent* on Sunday (21 June 1998), as cited in Haegeman & Ihsane 1999.

Several different characterizations of radical *pro*-drop have been offered in the literature, and it is worth considering whether any of these proposals could provide insight into the diary null subjects in typically non-*pro*-drop languages. Neeleman & Szendroi (2007) provide an overview of some existing accounts, including an analysis of radical *pro*-drop as a form of topic-drop. In section 1.5.2.5, I considered the topic-drop analysis of diary null subjects and found it wanting; Neeleman and Szendroi provide arguments against a general analysis of all radical *pro*-drop as topic-drop.

It has often been noted that radical *pro*-drop languages such as Chinese and Japanese not only lack rich agreement, but in fact have no agreement at all. Jaeggli & Safir (1989) maintain that *pro*-drop is related to agreement, and propose that subjects are only obligatory when there is *poor* agreement rather than rich agreement or no agreement at all. Since Chomsky (1981), subject-verb agreement has been characterised in terms of a specifier-head relation between the head of the inflectional node (I) and the specifier of the inflectional node [SPEC-I]. A head may contain grammatical features known as ϕ -features which typically relate to person, number and gender. When those features are unspecified, their values must be provided via specifier-head agreement with a subject. Pronouns, anaphors and R-expressions each have grammatical features, and so may license the agreement. *Pro*, however, lacks ϕ -features and so cannot act as a licenser. Speas (1994; 2006) relates agreement with *pro*-drop in an analysis based on the assumption that by the end of the syntactic derivation all ϕ -features must be specified. In languages with rich agreement, all ϕ -features are fully specified already, removing the need for a subject. In languages without agreement, there are no ϕ -features to be specified, and so no need for an overt subject. However, in languages with poor agreement, some, but not all, of the ϕ -features are specified, and so an overt subject is required to specify those remaining features. Whilst this seems an elegant way to capture the correlation between poor inflectional morphology and overt subjects, Neeleman and Szendroi point out that counter-examples to this paradigm exist: for instance, they note that Swedish and Afrikaans both have no agreement, but *pro*-drop, radical or otherwise, is not a grammatical option in these languages.

Tomioka (2003) offers an alternative approach. He makes the general claim that all languages which allow radical *pro*-drop also allow bare NP arguments. Null pronouns are therefore ‘simply the result of N-deletion/NP ellipsis without determiner stranding’ (p. 336). This seems an intriguing approach from the diary drop perspective, as the diary-style texts typically also display determiner omission. However, Neeleman

and Szendroi again provide counterexamples: for instance, Cheke Holo has obligatory determiners and yet radical *pro*-drop is possible.

Having found the existing proposals unsatisfactory, Neeleman and Szendroi propose a new analysis of radical *pro*-drop which links the availability of radical *pro*-drop with the morphology of the pronominal paradigm. In brief, radical *pro*-drop is licit if a language ‘has at least some agglutinating pronominal morphology’ (p. 678). In languages with agglutinating morphology, the various morphemes making up a word are clearly demarcated and can be easily differentiated. For example, in Japanese, the nominative case affix ‘ga’ may be attached to the masculine singular pronoun stem ‘kare’ to form ‘kare-ga’. By contrast, in languages with fusional morphology, such as English, the individual morphemes are not always readily distinguishable from the stem or from one another. This is evident if we compare the accusative masculine pronoun ‘him’ with its nominative counterpart ‘he’. According to Neeleman and Szendroi, languages in which ‘case on pronouns is fusional do not permit radical *pro*-drop’ (p. 679), although agreement-sensitive *pro*-drop may still be allowed in traditional *pro*-drop languages such as Italian and Spanish. It seems immediately clear that this does not offer a general solution to the issue of diary null subjects in English, as the pronominal morphology does not alter in the registers which allow diary-style nulls. When pronominals in the diary-style registers are overt, they take the standard, fusional forms. However, in the course of their analysis Neeleman and Szendroi offer a minimalist characterization of the empty argument slots in both classic and radical *pro*-drop languages which differs from the possibilities available in the Government and Binding framework.

According to the Government and Binding theory of empty categories, *pro* is a silent pronoun which sits in the subject position [Spec, IP] and is interpreted by virtue of the agreement features on the Inflectional head. However, this characterisation is not compatible with developments introduced with the minimalist program (Chomsky 1995). *Pro* cannot have inherently valued features, since its interpretation varies with the context. Because *pro* may be interpreted as a referential subject, its features must be derived at some point in the syntactic computation. In GB terms, *pro* gets its feature specification from the I-node. However, this solution violates the principle of Inclusiveness (Chomsky, 1995; Neeleman & Van de Koot, 2002), according to which ‘the properties of a terminal node are recoverable from the lexicon and...the properties of a non-terminal are recoverable from the structure it dominates’ (Neeleman &

Szendroi, 2007, p. 683). The subject position where *pro* sits, [Spec, IP], does not dominate the I node, and so the features from I cannot be copied to it.

Neeleman and Szendroi's solution to this problem is to propose that *pro* carries a full set of features and that *pro*-drop is simply the 'zero spell-out of regular pronouns' (p. 683) according to a spell-rule such as (111).

(111) [+p, -a] → ∅

This rule is seen as applying to both the classic *pro*-drop null subjects and the radical *pro*-drop cases, and is one of the assumptions on which their analysis of radical *pro*-drop is built. In their view, null arguments are 'regular pronouns that fail to be spelled out at PF, rather than instantiations of a special silent lexical item, *pro*' (p. 679). This brings us to the main point of interest for this thesis: the idea that, as Neeleman and Szendroi acknowledge, ultimately the realization or otherwise of a pronominal argument may be a pragmatic matter:

What circumstances require phonological realization of a pronominal is a matter of debate, but it is clear that under the present proposal contrasts between overt and covert pronouns must be attributed to pragmatic considerations (2007, p. 685).

Holmberg (2004; 2005) reaches a similar conclusion. Again, applying minimalist principles, his work on Finnish leads him to conclude that null subjects are specified for interpretable ϕ -features, and that 'the nullness is a phonological matter' (2004, p. 3). He goes on to propose that 'deletion is a phonological operation, but one which is dependent on a condition of processing, namely recoverability of the deletion,' with recoverability being possible via 'agreeing auxiliary, verb or adjective in conjunction with contextual cues' (2004, p. 10). Having removed the issue of pronoun realisation from the domain of narrow syntax and presented it as a phonological and processing matter, Holmberg considers whether 'null subjects are in principle available in any language' (ibid). This issue is considered in more detail in Holmberg (2005), where he suggests that typically non-null subject languages have a 'stricter "phonological" EPP condition that requires not only a filled [SPEC, IP], but also a pronounced [Spec, IP]' (2005, p. 557). This in turn raises the question of the contextual and processing conditions under which this strict condition might be relaxed in non-*pro*-drop languages containing the diary-style nulls. This is the starting point for my pragmatic analysis of

the diary-style null subject, and seems largely compatible with the ellipsis conclusion reached by Haegeman and Ihsane. From now on, I will therefore assume that the subject is a fully specified, syntactically active pronominal which is optionally deleted under certain conditions via a phonological rule. I will argue that since these conditions involve both pragmatic considerations and processing constraints, they are best investigated using a cognitively oriented pragmatic framework such as relevance theory.

1.7 Concluding remarks and the next steps

Syntax clearly has something to say about the distribution of the diary-style null subjects. However, as I have tried to show, the jury is still out as to the best way to formalise the syntactic constraints and patterns. Although I am tentatively adopting the proposal put forward in Holmberg (2005), my main aim is to address an issue which is acknowledged by all existing syntactic accounts, and may well be the one thing they all agree on: the need for certain contextual and pragmatic conditions to be met if a null-subject utterance is to be both interpretable and appropriate. This is a theme that runs through all the existing attempts to analyse the diary drop data. In early work on the subject, Haegeman (1990a) expresses the view that ‘the accessibility of the referent is not a matter of syntax properly speaking but a matter of pragmatics’ (1990a, p. 167). Neeleman and Szendroi acknowledge that ‘a full theory of *pro* drop requires an additional pragmatic component that governs the use of null pronouns in languages whose grammar allows them,’ (2007, p. 673); and according to Tomioka, ‘the descriptive content of the null NP is pragmatically retrieved’ (2003, p. 329).

Although the pragmatic factors governing the use of diary-style null subjects have not so far been considered in detail, several pragmatic proposals have been put forward about what is often seen as a related phenomenon: the use of null subjects in child language. In chapter 2, I will consider these null subjects, which occur systematically during a particular phase of language acquisition, as a possible analogue to the adult diary-style forms. I will look at both syntactic and pragmatic analyses of child null subject data, and consider the strengths and weakness of these accounts and the potential for extending them to the adult domain. I will then introduce the main principles of relevance theory (Sperber & Wilson, 1986/95; Carston, 2002a; Wilson & Sperber, 2004), and discuss how they might shed light on the null subject examples in child language and help to integrate the informal pragmatic observations found in the existing literature.

In chapter 3, I return to the assumption that the diary-style nulls are non-overt pronominals, and therefore function as referring expressions. To examine the pragmatics of reference in more detail, I consider two influential pragmatically-oriented accounts of referring expressions: Accessibility Theory (Ariel, 1988; 1990; 2001) and the Givenness Hierarchy (Gundel, Hedberg, & Zacharski, 1993; Gundel & Mulkern, 1998). Both accounts are intended to be compatible with the relevance theoretic framework, but both claim that relevance theory alone is not enough to explain the use and distribution of referring expressions, and must be supplemented by additional machinery. In chapter 4, I challenge this view, and propose a relevance-based approach to referring expressions which does not require any theory-external notions of the type proposed by Ariel or Gundel. My central claims are, first, that referring expressions encode conceptual and/or procedural meaning which may contribute to what is explicitly and/or implicitly communicated by an utterance, and second, that these meanings are specifiable without invoking any notions from Accessibility Theory or the Givenness hierarchy. I go on to develop these claims and apply them to further data in chapter 5. In chapter 6, I return to the diary null subject data, and show how it fits into my account of referring expressions. I argue that null subjects in English are not exceptional, but emerge as a natural consequence of the speaker's aiming to make her utterance optimally relevant. In chapter 7 I discuss some outstanding issues, and draw some general conclusions about the nature of the procedural/conceptual distinction in light of my analysis of null subjects in English.

Chapter 2: Null Subjects in Child English

2.1 Introduction

2.1.1 The null subject phase

Children learning English go through a phase in their linguistic development where they systematically produce sentences like those in (1)-(5).¹

(1) missing there.

(2) ride truck.

(3) bump my train.

(4) want go get it.

(5) read bear book.

This phase generally occurs when the child is between 20 and 25 months old, although exact timings vary from child to child. The same pattern is observed in other non-*pro*-drop languages including French, as in (6) and Danish, as in (7) and (8):

(6) a tout tout tout mangé .

‘has all all all eaten.’ (Hamann, Rizzi, & Frauenfelder, 1996)

(7) er ikke synd.

‘is not a pity.’

¹ Examples taken from Hyams (1986) quoting from the corpora of Bloom (1970) and Bloom, Lightbrown and Hood (1975)

(8) ikke køre traktor.

‘not drive tractor.’

(Hamann & Plunkett, 1997; 1998)

According to a Principles and Parameters approach to language acquisition, a child must set her *pro*-drop parameter either positively or negatively. If she is learning Spanish or Italian the parameter will be set positively, and if she is learning English, French or Danish the parameter will be set negatively. Wexler (1998) provides evidence for what he terms the Very Early Parameter setting hypothesis of language acquisition (VEPS), arguing that, ‘parameters that are set at the earliest observed stage (i.e. at the beginning of production of multi-word combinations, around 1;6) include...Null subject or not’ (1998, p. 29). If this is the case, we would not expect to find null subjects in the speech of a child acquiring a non-*pro*-drop language after the age of 18 months. From this perspective, the data in (1)-(8), produced at a later stage in development, seem problematic.

However, closer inspection reveals that the environments in which null subjects occur during the null subject phase do not pattern consistently with those of adult null subjects in *pro*-drop languages. As noted in 1.5.3, they look rather more like adult diary-style null subjects. Haegeman (1997, p. 234) goes so far as to describe the distributions of child null subjects and adult diary-style null subjects as ‘strikingly similar’.

In this chapter, I take a closer look at the acquisition data and consider some of the analyses that have been proposed to account for it. I will then outline the basic principles of relevance theory, and suggest that this cognitively-orientated framework offers an insightful perspective on the pragmatic aspects of subject drop in child language. Although most accounts of both adult and child null subjects emphasise the importance of pragmatics for understanding the null subject data, and some make informal suggestions as to how we might characterise this, I suggest that adopting a relevance-based approach has the advantage of allowing us to examine the data from the perspective of an independently motivated pragmatic framework. My discussion of the child data in this chapter will set the scene for my relevance-based account of referring expressions in general, and diary-style nulls in particular, which I will develop over chapters 3 to 7.

2.1.2 Competence or performance?

Existing treatments of null subjects in child language fall broadly into three camps: competence (syntactic) accounts (Guilfoyle, 1984; Hyams, 1986; 1992; Kazman, 1988; Guilfoyle & Noonan, 1989; Hyams & Wexler, 1993; Rizzi, 1994, 2005; Bromberg & Wexler, 1995), performance accounts (Bloom L. , 1970; Bloom P. , 1990; 1993; Gerken, 1991; Valian, Aubry, & Hoeffner, 1996), and discourse-pragmatic accounts (Greenfield & Smith, 1976; Allen, 2000; Schaeffer, 2005). I begin by considering the first two categories in more detail.

The distinction between the roles of competence and performance in language production and processing was first highlighted in the work of Noam Chomsky (1965), where competence is tacit knowledge of language, and performance is the practical use of language in concrete situations. Language acquisition involves acquiring a grammar that allows us not only to produce and understand utterances in a certain language, but also to make judgments of well-formedness for sentences we might never have heard before. These intuitions about sentence meaning and well-formedness are seen as reflections of our linguistic competence.

However, there are many external mental or physical factors which may affect our ability to utilize this internal grammar to produce appropriate utterances or to understand the utterances of others. These 'performance' factors (e.g. false starts, slips of the tongue and other types of mis-speaking) may vary across persons, times and situations. We may, for example, be likely to make more performance errors when we are tired, drunk, nervous or distracted. Such errors are not taken to reflect a deficit in our underlying linguistic competence, but are rather treated as on-line glitches in production or comprehension. There are also more general performance factors which affect or constrain our linguistic output across circumstances. Language is recursive, and a grammar should, in theory, generate indefinitely long sentences. However, working memory places limitations on how much information can be kept active for use in computation at any one time, and there are corresponding practical length restrictions on language production. Similarly, we may have difficulty processing certain structures, such as multiple embeddings, so that sentences such as (9) or (10), which are perfectly well-formed sentences according to the grammar, are not usually able to be processed by our performance systems, at least not without considerable conscious effort:

(9) Bricks bricks break break bricks.

(10) The dog the stick the fire burned beat bit the cat. (Pinker, 1994, p. 205)

This distinction between competence and performance leaves us with two possible lines of explanation for the child null subject phase. It could be the result of some deficit in the child's underlying grammar (as compared with the adult's) or it could be due to performance factors. Competence-based approaches start from the assumption that all children pass through a stage where their grammar allows null subjects. Utterances with null subjects are therefore fully well-formed according to their internal grammar. Then, as acquisition proceeds, this grammar switches, changes or matures in such a way that it converges with the adult grammar of the language being acquired (Hyams, 1986; 1992; Hyams & Wexler, 1993; Bromberg & Wexler, 1995). According to performance accounts, on the other hand, the child is acquiring the target language grammar from the start, and the omission of subjects is the result of some form of processing constraint, perhaps combined with pragmatic considerations (Bloom L., 1970; Bloom P., 1990; Gerken, 1991).

Proponents of both approaches have put forward theoretical and conceptual arguments designed to show that a competence or performance approach is preferable on general grounds. Empirical evidence has also been provided in support of one or other approach. I begin by considering the claimed advantages of a competence approach, and outline some of the specific analyses that fall into this category. We will see that there are similarities between these accounts and the accounts of adult diary nulls outlined in the last chapter.

2.2 Competence Accounts

2.2.1 Advantages of a competence approach to child nulls

Hyams and Wexler (1993) argue that competence accounts are superior to performance accounts on theoretical, conceptual and empirical grounds. Discussing the distributional properties of the child null subject, they observe an apparent asymmetry between subject and object drop during this stage in acquisition. Data from the CHILDES corpus (MacWhinney, 2000) suggests that during the null subject phase children 'drop specific subjects about twice as often' (1993, p. 428) as they drop specific objects. This statistic, they claim, is most easily accounted for under a competence model. In their

view, performance accounts have difficulty explaining why performance factors might specifically target subjects in this way. The grammatical nature of the competence accounts, on the other hand, makes it possible to deal with this asymmetry by specifying that ‘the option to drop a specific argument is available only to subjects’ (p. 428). Hyams and Wexler go on to discuss correlations between the child null subject stage and various other specific stages in the language acquisition process, arguing that the acquisition of inflection (Meisel, 1987), tense (Clahsen, 1986) and non sentence-external negation (Pierce, 1992) all correlate with the end of the null subject phase. Competence-based accounts, they claim, are best placed to explain such correlations, since these accounts work from the assumption that ‘the child is developing an adult grammatical system and that, depending on the precise structure of the system and the various interrelations that exist, each development may generate a wide range of effects’ (Hyams and Wexler, 1993, p. 425).

Given these general arguments in favour of a competence-based approach to the child null subject phenomenon, I will now consider the details of some specific grammatical accounts.

2.2.2 The *pro* hypothesis

At least superficially, the subjectless utterances produced by children during the null subject phase can look like the subjectless sentences of a *pro*-drop language such as Italian or Spanish. Hyams (1986) begins from this observation and argues that the child’s early grammar differs from the adult grammar in that it has the *pro*-drop parameter positively set. On this account ‘young children speak a language with the essential properties of an adult *pro*-drop language like Italian, and the early grammar of English...represents a MIS-SETTING along a specific parameter of UG’ (Hyams, 1996, p. 94). This approach differs slightly from the *pro*-drop accounts of the diary nulls that we saw in the previous chapter. The child need only switch her setting once as she moves from a *pro*-drop setting to the non-*pro*-drop of the target language, whereas the adult must switch back and forth between the two settings as she moves between dialects. Hyams suggests that in its initial state the *pro*-drop parameter is positively set, and children learning English produce these subjectless sentences up until the point when the parameter is reset to negative. She also considers what might trigger such a resetting, and outlines the kind of evidence the child would need in order to make this change.

On Hyams' account, when the child hears a well-formed sentence of English that she is unable to generate with her current *pro*-drop grammar, this will provide a piece of evidence against a positive setting of the *pro*-drop parameter, and should be enough to trigger the switch. Such evidence would be provided by expletives. Given the Avoid Pronoun Principle², a language in which subjects are optional would always avoid the use of expletives, and indeed, expletives are not found in *pro*-drop languages. Thus, the presence of an expletive in the input to the English-speaking child should be enough to show that overt subjects are obligatory in subject position in the target language. According to Hyams, children begin to use expletives at around the time that they stop producing subjectless sentences, and she concludes that they therefore provide the necessary trigger. In sum, Hyams' 1986 analysis attempts to assimilate the child null subject with the null subject in languages such as Italian and Spanish.

However, this *pro*-drop analysis faces several objections, some of which were discussed in chapter 1 in relation to the diary-style nulls in adult language. Hyams (1992) revisits the data and identifies empirical, conceptual and logical problems with her earlier account. Valian (1990) also provides experimental data suggesting that the predictions of a *pro*-drop analysis are not borne out.

The most striking objection to a *pro*-drop analysis has to do with structural differences between the distribution of the child null subject and adult *pro*. If the child at this stage of development has her *pro*-drop parameter positively set, then we would expect to find the child null subjects occurring in the same environments as in adult *pro*-drop languages. According to Rizzi (2005), however, the child null subject is found very rarely, if at all, in subordinate clauses, whereas in true *pro*-drop languages such as Italian, subordinate clauses are a licit environment for *pro*. Thus, sentences such as (11) are perfectly acceptable in Italian and are found in the speech of both adults and children. By contrast, equivalent sentences, such as (12), are unattested in child English.

(11) Ho detto che _andava a casa.

(12) *I said that _ went home.

(Hyams 1996)

² See Chomsky (1981) and Hyams (1986), along with chapter 6, section 6.2.2 for further discussion of the Avoid Pronoun Principle.

As Rizzi (1994) points out, this evidence may be less significant than it appears, since children do not produce genuine subordinate clauses until their mean length of utterance (MLU) has gone beyond that usually associated with the null subject phase. However, he does give examples of occasional utterances of this sort that have been attested in this phase, for example (13).

(13) _ know what I maked.

Here, the subject in the main clause, but not the subordinate clause, is omitted. Rizzi treats this as evidence that null subjects in child English are restricted to main clauses (although he acknowledges that further investigation is needed before firm conclusions can be drawn). Empirically, then, it seems that in the early language examples, we are dealing with a phenomenon that differs significantly from adult *pro*-drop.

As noted above, all children drop subjects. If this is because all children in this phase have the same, *pro*-drop grammar, then we would expect the ratio of overt to null subjects to be constant cross-linguistically. Valian (1991) reports on a study that reveals that American children learning English produce at least twice as many overt subjects as Italian children. This distributional difference suggests that there is some difference between the two developing grammars, and that the underlying cause of the subjectless sentences may not necessarily be the same.

Conceptually, the idea of a mis-set parameter is also troublesome. There is no evidence that such mis-setting occurs with other parameters, and, indeed, the evidence rather suggests that parameters are set very early in the child's language development (Hyams 1996; Wexler 1998). Logically too, Hyams' 1986 approach is somewhat problematic. Valian (1990) discusses some objections to the notion of a single-value default parameter. If the parameter has a default setting, then the question arises of how the child comes to reset it. If, as Hyams suggests, the default for *pro*-drop is a positive setting, then the presence of sentences with subjects is not enough to invalidate this initial setting, since sentences with subjects are optional in null subject languages. It follows that the set of sentences with overt subjects is a subset of the set of possible sentences in a null subject language, and exposure to sentences with subjects will not on its own be enough to prove that subjects are obligatory. There is also the added complication that a child acquiring English or another non-*pro*-drop language is likely to hear subjectless sentences as part of the input. As discussed in chapter 1, subjects may at times remain non-overt in sentences of casual spoken English as well as in the

diary-style registers. Yet according to the *pro*-drop account, children must somehow manage to reset their parameter, and so end up with a non-*pro*-drop grammar despite this contradictory input.

As discussed above, Hyams suggests that expletive use is specific to non-null subject languages. However, this too proves to be problematic. If the child's grammar is parametrically set to *pro*-drop, then she will be unable to construct a representation for expletives when they do occur. They will, in effect, be unanalysable and, therefore, filtered out. Alternatively, as Valian suggests, the child with a *pro*-drop grammar may assign a referential interpretation to any expletive pronouns she hears. Nor is this the only problem for the expletive explanation of parameter resetting. Valian cites empirical data that bring into question Hyams' basic assumptions about the distribution of expletives in child language. If expletives are the trigger for the parameter resetting, we would predict that the child should only start producing them at the end of the null subject phase. Although the overall frequency of sentences with expletives is low in child language, contrary to Hyams' (1986) predictions, Valian found instances of their use across the age and MLU range studied. This included children who were still firmly in the null subject stage.

In Hyams (1992) the author herself outlines some problems with her earlier analysis. In non-*pro*-drop languages such as Spanish and Italian, it is often claimed that the rich inflectional systems license identification of the subject. Children acquiring these richly inflected languages master the inflectional systems at a very early stage. Whilst the inflectional systems of non-*pro*-drop languages such as English and French are considerably poorer than their *pro*-drop cousins, children acquiring them do not master the inflectional paradigms until well after they have begun to produce null subjects. As a result, the null subjects of children acquiring non-*pro*-drop languages are 'unidentified' (Hyams, 1992, p. 253). There is no inflectional information to identify the missing subject.

Hyams therefore moves away from her earlier focus on inflection, and presents an adjusted account which treats morphological uniformity as the crucial factor determining whether a language allows null subjects or not. As noted above in chapter 1, whilst *pro*-drop languages such as Italian and Spanish have rich inflectional systems, there are also languages, such as Chinese, which have no inflection but nonetheless allow null subjects. To incorporate this data, Jaeggli and Safir (1989) propose that 'null subjects are permitted in all and only those languages which have morphologically uniform inflectional paradigms' (Hyams 1992 p. 253). Accordingly, Hyams suggests

that during the null subject stage, English speaking children analyse their language as morphologically uniform (and more particularly, as non-inflectional). This leads to two predictions: first, that inflection will be omitted during the null subject stage, and second that when inflection is acquired, the child will cease to use null subjects. Hyams claims that these predictions are borne out by the data. Although this modified account addresses the problems that Hyams herself identified with her 1986 approach, many of the other objections remain. It is still unclear how we might explain the distributional difference between child null subjects and adult *pro*-drop, and how we might account for the cross-linguistic differences in the frequency of child nulls. The conceptual issues relating to the notion of a mis-set parameter also remain.

2.2.3 The VP hypothesis

Valian (1991) uses data from American and Italian children to test the predictions of both the pro-hypothesis and an alternative grammatical approach which she terms the VP-hypothesis. This competence-based analysis (Guilfoyle, 1984; Kazman, 1988; Guilfoyle & Noonan, 1989) proposes that the child's immature grammar generates only a VP, with no inflectional phrase (INFL) or complementizer phrase (COMP). In the mature, adult grammar, the need to check case means that the Spec of VP must be filled with an NP that can move to become the subject of the INFL phrase. Since the child's immature grammar provides no INFL phrase, there are no sentential subjects. Instead, when subjects do occur, they are located in the Spec-VP position. As this position is optional, so too are subjects. This account predicts that the language of a child in the null subject phase will also lack infinitival 'to', modals and nominative case marking.

Valian (1991) tests these predictions against the American and Italian data, and finds them not to be borne out. The subjects produced by the American children 'look like real subjects rather than VP subjects, because the children consistently used nominatively case-marked pronouns in subject positions' (1991, p. 76). Furthermore, American children produce twice as many subjects as Italian children, and they also contrast with the Italian speakers in that the majority of their subjects are pronouns. Although the American and Italian data in general contained few modals, children acquiring American English produced more, and no correlation was found between the onset of modal use and the loss of the subjectless option.

Valian notes, however, that the predictions of the VP hypothesis do seem to hold true for the very youngest child in her data set, who also had the lowest MLU. She therefore entertains the possibility that at this very early stage the child does indeed

have a VP grammar. However, since the use of subjectless sentences persists well past this stage, other or further explanations are needed for the later data.

2.2.4 Topic drop

As noted above, in certain languages, including Chinese, Korean and Japanese, null subjects are licit despite the impoverished inflectional systems of these languages. It has been argued that what distinguishes such languages is not the *pro*-drop parameter, but a parameter that allows zero topics in so-called ‘discourse-orientated’ languages (Huang, 1984, p. 551). These languages also differ from traditional *pro*-drop languages in that they allow the omission of objects as frequently as the omission of subjects. This is to be expected if the null element is a topicalised constituent, rather than an unpronounced subject pronoun. Topic-drop can also be found in some languages more traditionally thought of as non-null subject. In Dutch, for instance, any constituent may be topicalised, and hence appear in first position: [Spec CP]. In certain pragmatic conditions these topics may then be dropped, as in (14), where the name of the film in question (Rainman) might be seen as being topicalised and then dropped.

- (14) a. Ga je mee naar Rainman vanavond?
Go you to Rainman tonight?
- b. Heb ik al gezien.
Have I already seen.

Haan and Tuijumen (1988) argue that Dutch children have a process of Topic-drop. Both subjects and objects may be dropped when in topic position, but subjects are hardly ever dropped when not in first position. Similar accounts have been suggested for the English data. These topic-drop approaches hypothesize that during the null subject phase, children acquiring English have a topic-drop grammar (Hyams & Jaeggli, 1988; Hyams & Wexler, 1993; Bromberg & Wexler, 1995). As discussed in chapter 1, movement of a constituent to topic position is a commonly attested phenomenon in English, with examples such as (15).

(15) a. John, I spoke to yesterday.

b. CP [John_i [I spoke to t_i yesterday]]

Furthermore, Bromberg and Wexler (1995) suggest that adult English may have a limited capacity for topic drop, which allows subjects to be dropped when they are topics, as in (16) (1995, p. 243).

(16) a. What happened to Mary?

b. ___ went away for a while.

However, unlike Dutch children, English speaking children do not drop objects. Their topic-drop is restricted to subjects.

Hyams and Wexler (1993) attempt to explain these distributional facts whilst maintaining a topic-drop analysis of the child null subjects of finite verbs.³ According to their version of the topic-drop analysis, a constituent may be dropped if it is outside the VP. Subjects are already outside the VP, and will therefore be candidates for omission. According to Hyams and Wexler (1993), there is good evidence that Dutch speaking children can topicalise objects as well as subjects (Wexler, 1991; Hyams, 1994), and so both may be subject to omission. In contrast, the fact that English has a ‘much less robust pattern of topicalization’ (Hyams & Wexler, 1993, p. 430), leads Hyams and Wexler to conclude that, at this stage, English speaking children do not have a process of topicalization. As a result, objects will never move to a position outside the VP, and so will not be dropped.

According to Hyams and Wexler, the cross-linguistic variation in frequency of null subjects discussed by Valian (1991) is also predicted on this topic-drop account. Italian-speaking children have a *pro*-drop grammar where the empty subjects are identified via rich inflection. All subjects are therefore suitable candidates for dropping. In the case of the English-speaking children, however, not every subject will be a topic, and so the overall rate of omission will be lower.

³ Bromberg and Wexler (1995) suggest that two kinds of null subjects are represented in child language: those resulting from a topic-drop process and those which are licensed by a non-finite verb and are related to an optional infinitive stage in the child’s grammar.

Bromberg and Wexler (1995) argue that the absence of null subjects in tensed *wh*-questions provides further evidence for a topic-drop analysis⁴. Since topics and *wh*-words target the same tree node in the syntactic structure, they cannot co-occur in the same sentence. If the absent nulls are topics, then we should expect to find no instances of *wh*-questions amongst the null subject data.

Although the topic-drop analysis seems to account for many of the child null subject distributional facts, several questions remain unanswered. Topic-drop accounts do not explain why maturing speakers lose the option to use the topic-drop construction that they so freely used as children. Indeed, Hyams and Wexler acknowledge in a footnote that their analysis does not address either this issue or the issue of how the phenomenon relates to the adult language in general. This seems less than ideal for a theory of child language, and leaves as many questions unanswered as it addresses.

2.2.5 Truncation

The so-called Truncation account of the child null subject was developed by Rizzi (1994). It is this analysis that forms the basis for Haegeman's (2000) analysis of adult diary-drop outlined in chapter 1.

Recall that, according to Rizzi's amended identification condition (Rizzi, 1994), if the empty category is in the highest position in the structure and there is therefore nowhere for an antecedent to sit, then the need for the antecedent is waived, and, instead, identification takes place via the discourse. Rizzi proposes that, in the child's grammar, the empty category is indeed in the highest position because the child does not obligatorily project to the CP level (see chapter 1, section 1.6.1). When a CP is not projected, null subjects are licensed via Rizzi's amended ECP. In sum, the child's grammar lacks a rule specifying that the root must project to CP (ROOT=CP), and until this is acquired as an obligatory condition, null subjects may occur. We therefore see the same process as in the parallel adult account, but with a different motivation. In the adult cases, it is the special circumstances of certain registers which block the projection to CP level, whereas in the child cases, it is the immature state of the child's grammar which stops this projection from being mandatory.

This account clearly predicts that null subjects will not occur in sentences where *wh*-movement has taken place. For *wh*-preposing to take place, a CP level must be projected, which means that potential antecedent site will be available and, according to

⁴ See section 2.2.5 below for further discussion of the distribution of null subjects with *wh*-preposing

the identification clause of the ECP, must be used. Rizzi claims that this is indeed the case and that child null subjects and *wh*-preposing do not co-occur. However, Bromberg and Wexler (1995) present data from the CHILDES corpus which contradicts this claim. They find that ‘null subjects are abundantly present in *wh*-questions’⁵, citing minimal pairs such as (17) and (18) as evidence:

(17) Where go?

(18) Where dis go?

So again we find ourselves with a grammatical, competence-based account which deals with certain aspects of the phenomenon, but is left wanting when further data is considered.

2.2.6 Summary

In this section, I have given an overview of some of the major competence accounts of null subjects in child language. According to some of these, child nulls are the result of the child’s having a *pro*-drop or topic-drop grammar during a specific stage in the acquisition process. Alternative accounts work from the assumption that the child’s syntactic representation is truncated. Whilst I have tried to show that there are problems with the specific details of these accounts, this does not necessarily invalidate the general claim (Hyams and Wexler 1993) that competence accounts in general are best placed to deal with the child null subject data. Hyams and Wexler specifically argue that performance accounts are unable to explain the subject-object asymmetry and the association between the null subject phase and ‘a range of other theoretically related grammatical properties’ (1993, p. 452). With this in mind, in the next section I will outline the arguments in favour of a general performance-based account of the data, and consider some of the performance analyses that have been proposed.

⁵ Note that Bromberg and Wexler’s own account combines a topic-drop analysis (for tensed clauses) with an analysis which links null subjects to the ‘optional infinitive’ phase. They conclude that for *wh*-questions: ‘tensed sentences never have a null subject and untensed sentences do a significant portion of the time’ (1995, p. 239).

2.3 Performance Accounts

2.3.1 General advantages of performance-based accounts

The previous sections considered possible ways of treating the child null subject stage as resulting from either an immature grammar or a mis-setting of a parameter during the acquisition process. The alternative is to analyse the child's output as a matter of performance. The assumption behind this approach is that the child's grammar is not qualitatively different from the adult's, but that the output is affected by factors relating to the child's ability in production.

General evidence in support of performance analyses is provided by Lois Bloom (1970), who argues that the child knows more about the adult grammar than is reflected in her own utterances. Bloom reports an experiment by Shipley, Smith and Gleitman (1969) which reveals that children, including those in the 'telegraphic' phase, respond more readily to full, well-formed commands than to commands expressed in a telegraphic style similar to their own. This suggests that children know more about the rules and structure of language than is superficially evident from their own utterances, and that we should take performance factors seriously when analysing child language production.

There are several other empirical advantages to a performance-based account. For instance, children during the null subject stage do not only omit subjects. As in the adult diary-style registers, other categories, including determiners, verbs, auxiliaries and prepositions, are liable to omission during this stage. Ideally, an account should shed some light on these further omissions. As Bloom notes, 'no one has proposed a parameter of subject-determiner drop' (1990, p. 727).

Bloom (1990) and Valian (1991) both provide evidence that the change from the subject drop stage to adult-like subject use is gradual. Competence accounts would seem to predict a more abrupt change as the child re-sets the appropriate parameter or acquires the necessary grammatical structure. A performance account incorporating reference to the child's developing cognitive systems seems better suited to dealing with gradual change.

It is worth emphasising here that competence and performance analyses should not be seen as mutually exclusive. Indeed, many competence-based accounts of child null subjects acknowledge at least some role for pragmatic and performance factors. Hyams and Wexler (1993, p. 452) note that 'It is a trivial observation that children are limited in their productive abilities', whilst for Rizzi (2005, p. 24), language is

‘grammatically based, but performance driven’. Even in accounts that do not explicitly acknowledge the role of performance factors, there is clear evidence of exceptions to the grammatical rules or patterns. For example, Hyams and Wexler (1993, p. 428) explain the subject/object asymmetry via a grammatical model in which ‘the option to drop a specific argument is available only for subjects’. Yet their own data reveal instances of object drop, albeit at a much lower rate than subject drop. Similarly, Hyams (1996) concludes that, based on 94-99% of cases, modals occur ‘almost exclusively’ with overt subjects. In both cases it is claimed that a competence account provides the best fit with the data. However, even this best fit leaves a number of exceptions to be dealt with, and it is not clear how is this to be done if not in terms of performance.

The grammatical accounts considered above mostly deal in absolutes. A certain utterance is or is not well-formed according to a particular grammar. Such accounts, in some sense, abstract away from practical production and comprehension considerations. As Paul Bloom (1990) points out, performance accounts, by contrast, deal with a ‘tendency’ rather than an absolute. Whilst performance factors play some role in each of the processing analyses of child null subjects, in this section, I concentrate on accounts which place performance factors at the heart of their analyses.

Most performance-based accounts focus on processing limitations in the child. The idea is that omissions in production occur because the child’s capabilities are overloaded: some form of constraint on processing ability in the developing stage, combined with complexity in certain sentences, leads to reductions and omissions in the output. I begin by considering the earliest of these accounts, proposed by Lois Bloom, which focuses on the complexity added by extra sentence length. I then move on to other accounts which focus on factors such as VP length and metrical complexity.

Processing accounts also rely, to varying degrees, on pragmatic factors to explain why certain constituents are more vulnerable to omission than others when processing capacity is limited. As a result, there is some overlap between the performance-based and pragmatics-based accounts of the data, although none offers a fully integrated approach. I consider discourse-pragmatics-based accounts separately in section 2.4. In section 2.5, I will argue that by taking relevance theory as a pragmatic framework, we can develop an integrated approach which builds on the strengths of both kinds of account. The implications of relevance theory for the production and interpretation of subject noun phrases in general will be explored in more detail in chapters 3-5. In chapter 6, I return to the issue of child null subjects in light of my analysis.

2.3.2 Processing accounts

2.3.2.1 Sentential complexity: a general processing approach

The earliest performance-based account was that provided by Lois Bloom (1970), who analyses the speech of three children in the null subject stage and claims that complex sentences, combined with the child's cognitive limitations, lead to omissions in production output. Bloom categorises the speech of the children as 'telegraphic', likening it to the utterances produced by an adult 'who is under pressure to be brief' (1970, p. 139).

Considering which factors might cause the child's output to resemble the 'telegraphic' utterances of adults, Bloom finds that the length and, more specifically, the complexity of an utterance contribute to the likelihood that the output will be abbreviated. Between the ages of 20-23 months, the child is more or less limited to two word utterances. Where a longer utterance would have been required by the adult grammar, the child omits one or more of the words. Initially, this may look like a simple length constraint. However, Bloom suggests that what is at work here is rather a 'cognitive limitation in handling structural complexity' (1970, p. 165). To investigate this hypothesis, she considers the effect of negation on the child's utterances, on the assumption that negation increases sentential complexity, and she provides an in-depth analysis of its use in the speech of children during the telegraphic phase. She finds that, in sentences with negation, the probability that the utterance will be in some way reduced is higher than in sentences of comparable length without negation, and concludes, more generally, that the omissions in child language are due to a limitation on the linguistic and logical complexity with which the developing child can cope. As the complexity of a sentence increases, so do the chances that 'something had to give in its production' (1970, p. 165).

Under Bloom's analysis, both linguistic and cognitive factors play a role and may interact to determine both when an omission is likely to occur and what type of constituent is most likely to be omitted. Whilst a linguistic factor such as unfamiliar vocabulary or a logical factor such as negation may constrain the child's ability, Bloom also notes that cognitive constraints, such as reduced memory span, make the child's task still more complicated. The relation between memory and language acquisition is considered in more detail by Olson (1973). Rather than seeing the child's limited memory span as an underlying cause of the shorter utterances, he presents both phenomena as symptoms of the same process: it is the child's as yet underdeveloped

abilities to ‘recode, encode, to plan and monitor, to integrate and unitize’ (p. 153) which underlie both limitations.⁶

Whereas the grammatical, competence-based accounts concentrate on the omission of subjects, Bloom’s account considers the overall reduced nature of the speech of children during this phase. Her evidence suggests that subjects are not the only elements to ‘give’ under the pressure of cognitive constraints. Omissions are not random, and Bloom draws on the findings from previous experiments (Brown & Fraser, 1963; Brown & Bellugi, 1964) to argue that the omissions are in fact both predictable and systematic. Words that are persistently retained in the surface utterances tend to be ‘contentives’ (nouns, verbs and adjectives) whilst those which are omitted tend to be ‘functors’ (articles, prepositions, auxiliaries and pronouns). Here we see a parallel with omissions in the adult diary style contexts discussed in chapter 1. Bloom notes that ‘functors’ tend to be ‘weakly stressed, carry little information value [and] are most predictable’ (1970, p. 140). In this way, she seems to be making a move towards incorporating pragmatic factors into her processing-based account. In sum, Bloom’s account suggests that complexity resulting from syntactic or logical structure or from unfamiliar vocabulary may overload the child’s processing capacities. Given these constraints, the child produces those parts of an utterance which ‘carry the most information and are least predictable’ (1970, p. 140). Whilst this is a fairly general analysis of child language during the telegraphic phase, Bloom’s ideas are developed in more detail by subsequent performance accounts.

2.3.2.2 VP length and rightward complexity

Paul Bloom (1990) builds on Lois Bloom’s performance account, and presents results from a further study in its support. He focuses on the structure of the sentence, citing three strands of empirical evidence which favour a processing account over a competence approach. First, like Lois Bloom, he points out that various different types of constituent are omitted alongside subjects during this phase. The correlation between null subjects and other null constituents is not predicted on a grammar-only account. Second, he cites evidence from experiments by Brown and Fraser (1963) which show that constituents are omitted in imitated speech in the same way, and to the same degree, as in spontaneous speech. This suggests that the crucial factor is not so much a grammatical problem or difference from adult speech, but that the child may simply

⁶ Olson also considers the child’s ‘highly egocentric’ (p.155) view of the world as a factor contributing to the frequency of abbreviated utterances.

have ‘a general difficulty in producing long strings of words’ (Bloom P. , 1990, p. 492). Finally, he reports data from Mazuka et al (1986), which reveal that some children phonologically reduce their subjects to a schwa, rather than omitting them altogether. These data are hard to explain on a competence account where the child’s grammar either allows null subjects or doesn’t. By contrast, they are predictable on a performance account where the child knows a subject is necessary but lacks the necessary resources to consistently realise it phonetically.

Paul Bloom adopts Lois Bloom’s proposal about the significance of sentence length and tests it with specific attention to VP length. Counting the number of words from the verb position to the end of the sentence, he confirms the predictions of the performance-based account by finding a significant difference in the length of VPs in sentences with and without subjects. When a subject was overtly realised, the VP tended to be shorter than when it was omitted. As Bloom himself points out, there is an alternative way of accounting for this finding in purely pragmatic terms, since extra length in the VPs is likely to provide extra contextual information, which should make the subject referent easier to identify, and therefore more likely to be omitted. Bloom tests the predictions of this pragmatic account against those of the processing approach by comparing the length of overt subjects produced by children in this stage with the length of VPs in the sentences in which they occur. Whilst a pragmatic account (as he envisages it) predicts that the length of the overt subjects should have no effect on VP length, the processing account predicts that the extra effort involved in processing a complex subject will result in a shorter VP than for a simple subject⁷. Bloom found that his processing hypothesis was confirmed: there was a gradual decrease in the length of VP as the length of the subject increased.

In Bloom’s view, a purely pragmatic, discourse-orientated story cannot fully account for the null subject phenomenon. He quotes findings from Goldin-Meadow & Mylander (1984) which suggest that the likelihood of omission of a particular element cannot be predicted in terms of whether it carries old or new information. However, he does suggest that the distinction between old and new information might help to explain the observed object/subject asymmetry. Objects are more likely to convey ‘new’ information than subjects, which are more likely to convey ‘given’ information. His suggestion is that ‘given’ subjects are more likely to be omitted when processing capacity is overloaded. Alongside this suggestion, he also provides an alternative

⁷ As children in this range rarely produce complex subjects, in practice, Bloom compared the use of the unambiguous pronouns ‘I’ and ‘You’ with non-pronoun subjects.

explanation for the subject-object asymmetry, based on what he sees as a general linguistic bias to 'save the heaviest for last' (1990, p. 501). The idea behind this proposal is that processing load is 'proportional to the number of yet-to-be expanded nodes' (p. 501) in the syntactic representation. It follows that sentence initial subjects are associated with greatest number of unexpanded nodes, and therefore impose the largest processing load. As a consequence, subjects should be omitted more frequently than objects occurring later in the sentence.

Bloom ends by noting that his processing account is not incompatible with more competence-based mis-set parameter accounts. However, he points out that a performance approach explains many of the data which motivated the grammatical accounts, and that much of the reason for positing a competence account in the first place is therefore lost. He also suggests that the problems posed by competence accounts, such as the issue of how the grammar changes, are substantial drawbacks when weighed against performance accounts

2.3.2.3 A metrical approach

Taking a slightly different angle, LouAnn Gerken (1991) recapitulates some of the problems with competence-based accounts. She goes on to assess a variety of performance approaches, and argues that they all fail to capture the generalization that if the omitted elements were overtly realised, they would tend to be weakly stressed. Her alternative analysis is based on the hypothesis that children tend to 'omit the weak syllables from iambic [weak-strong] feet' (1991, p. 437). She notes, first, that children are more likely to omit a weak syllable at the beginning of a word than in word-final position. For example, both 'giraffe' and 'monkey' are two syllable words, but 'giraffe' has the main stress on the second syllable, whereas in 'monkey' the main stress comes first. According to her data, it is much more likely that we will find 'giraffe' being reduced to 'raffe' than 'monkey' being reduced to 'mon'. Gerken claims that this same tendency towards deletion of weak initial syllables applies at the level of sentences as well as individual words, so that subjects, as weakly stressed sentence initial elements, are most vulnerable to omission.

Gerken places her account firmly in the processing limitation camp, claiming that her metrical hypothesis 'provides a mechanism by which some sentential elements are omitted when sentential complexity becomes too great' (p. 443). She goes on to suggest that the hypothesis may also provide 'a measure of the sentence complexity itself', since sentences with pronoun objects differ in metrical complexity from those with lexical NP

objects. Her approach would therefore seem to satisfy many of our requirements for a comprehensive analysis of the phenomenon. It can account for the object/subject asymmetry, it appears to be supported by experimental data, it accounts for at least one aspect of the complexity which presents problems for the child, and it offers a systematic mechanism for reducing the processing requirements.

However, on closer inspection, Gerken's account faces some problems. Hamann and Plunkett (1998) point out that, although her approach works well for the English data, it does not hold cross-linguistically. For example, Gerken's explanation for the subject/object asymmetry is that 'children omit weak syllables from iambic feet [(he) COMES] more often than from trochaic feet [HIT him]' (Hamann & Plunkett, 1998, p. 39). In French, subject and object pronouns are clitics which usually occur before the finite verb, and object clitics may occur in iambic or trochaic feet. On Gerken's account, object clitics in trochaic feet should be omitted less often than those in iambic feet. However, this is not the case, since 'object clitics fail to occur in either trochaic or iambic feet' (Hamann & Plunkett, 1998, p. 39). Hamann and Plunkett also point out that Gerken's account predicts that the sentence and word level omissions will occur at the same stage in development, and they present evidence from Danish showing that this is not necessarily the case cross-linguistically. In sum, whilst offering a fairly convincing account of the English data, Gerken's metrical account does not hold up to cross-linguistic scrutiny.

2.3.3 Testing the predictions

Both the competence and performance-based accounts make fairly robust predictions about what we should expect to find in the language of children in the null subject phase, and how their language use might pattern. Valian, Hoeffner & Aubry (1996) test these predictions using data from two groups of children: one inside the MLU range associated with null subjects, and one outside. Using an imitation method where the children were asked to repeat what the experimenter had just said, they measured a variety of different factors, and their findings are summarized as follows. First, MLU and age were found to correlate. Second, pronominals were imitated less often than full lexical subjects. Third, sentence length was a factor in subject use for the low MLU group, but not for the group with high MLUs: children in the low MLU group imitated subjects in short sentences more often than in long sentences. This correlation between sentence length and subject omission is not expected on the competence-based accounts. Fourth, introduction of a topic affected both MLU groups to the same degree, and use of

expletives was the same across both groups. Again, this is not expected on the competence-based accounts, which predict that children outside the null subject MLU range should produce expletives more often than those within it. Fifth, all children used some functional elements, and no link was found between production of inflection and production of subjects. Finally, expletives were imitated less often than referential pronouns.

Each of these conclusions supports the hypothesis that the child subject omissions are the result of performance factors rather than a competence deficit. The results pattern more closely with a performance account where the child's cognitive and processing capabilities gradually increase, than with an account where the lower MLU group have different grammatical competence. The finding that expletives were imitated less often than referential pronouns is problematic for a purely metrical analysis, since they are both 'low-stress sentence-initial syllables' (Valian, Aubry, & Hoeffner, 1996, p. 161).

The authors conclude that the two groups they studied have the same competence, and that even very young, very low MLU children understand that English requires subjects. Like Paul Bloom, they found a correlation between subject use and both sentence length and VP length, and they conclude that a performance account explains the data more comprehensively than the grammatical accounts.

2.3.4 Comparing approaches

I return again to the claim made by Hyams and Wexler (1993) that performance accounts are unable to explain either the subject/object asymmetry found in omissions in child language, or the association between the end of the child null subject stage and the acquisition of other theoretically associated grammatical categories (e.g. inflection).

In the previous section I outlined the findings of Valian, Aubry and Hoeffner (1996) which cast doubt on the validity of the second claim. There does not seem to be a clear link between inflection or other functional categories and the production of subjects. How, then, might a performance account deal with the objections concerning the subject/object asymmetry?

As we have seen, Bloom (1990) presents possible pragmatic and processing explanations for the subject/object asymmetry. He develops these arguments in his 1993 reply to Hyams & Wexler (1993). The asymmetry, he argues, is to be expected in both adults and children, given the pragmatic and processing differences between subjects and objects. Not only is there a preference to 'save the heaviest for last', as

described above, but there is also evidence that adults are more likely to hesitate at the beginning of an utterance, suggesting that there is an increased processing load at this point. Assuming that this is the case, and that the child is therefore functioning under processing limitations, Bloom suggests that pragmatically redundant information is more likely to be omitted, and that subjects are more likely than objects to carry redundant information. To support this claim, he provides evidence that subjects are more often pronominalised than objects, and that nonpronominal subjects are significantly shorter than nonpronominal objects. These are aspects of a subject-object asymmetry that cannot be handled in a purely grammatical account.

In sum, the apparent advantages of a competence account can be either dismissed as unfounded or carried across into a performance account. By contrast, it remains unclear how the advantages of a performance account might be carried across to a competence account. In particular, the omission of other elements alongside subjects during the null subject stage and the gradual change to adult-like production are both hugely problematic for grammatical competence-based accounts.

2.4 Discourse pragmatic approaches: two types of informativeness

2.4.1 The role of pragmatics

In general, pragmatic accounts of child null subjects have received much less attention and are far less developed than their processing-focused cousins. However, most processing theorists acknowledge at least some role for pragmatics. As noted above, Paul Bloom (1993) suggests that, although the child's ability to perform is affected by processing limitations, children 'also have some control over what to omit', and they choose to omit 'pragmatically redundant material' (1993, p. 726). But he does not develop this suggestion much further, and the emphasis in processing accounts remains firmly placed on processing load as the most significant factor. In this section, I turn to discourse pragmatic accounts which place the main emphasis on pragmatic rather than processing factors, and consider two approaches which appeal to different notions of 'informativeness' in dealing with child null subjects.

2.4.2 Informativeness 1: Greenfield & Smith (1976)

Greenfield and Smith (1976) offer a pragmatic perspective on the null subject stage. They base their approach on a notion of informativeness, in what they call 'the

information-theory sense of uncertainty' (1976, p. 184)⁸, and claim that the most uncertain or most informative elements in the utterance content will be linguistically encoded, whilst less informationally rich elements are omitted⁹. In their view, an agent in subject position is the most obvious of the 'situational elements that can be taken for granted' (p. 108), and this is why subjects are so frequently dropped. Their theory is based on a study of several children's speech, which reveals that subjects are 'expressed infrequently,' especially in single word utterances and in sentences in which the children are referring to themselves. They find, in particular, that the AGENT concept is only overtly expressed in contexts where there is some uncertainty about who the referent is: these might involve a change in agent, or a conflict between one or more agents. Greenfield and Smith summarise their account as involving a 'pragmatic presupposition' that whatever the child can assume or take for granted is not overtly expressed. They go on to draw parallels between the null subject stage in the child's language development and what they call 'telegraphic ellipsis' in adult conversation, claiming that the child in the null subject stage is acquiring the ability to combine linguistic and non-linguistic information.

One problem with this account is that, although it provides a possible basis for choosing which items are to be omitted and which overtly expressed, it does not explain why such omission is necessary. Another is that Greenfield and Smith are more concerned with one-word utterances in general than with subjectless sentences in particular. As a result, they deal only briefly with issues such as the subject/object asymmetry found in child language, commenting merely that the child may 'be egocentric in taking more elements of his own perspective for granted' (1976, p. 195), and that we might therefore expect them to show a preference for subject omission.

Despite these issues, Greenfield and Smith offer a new perspective on child null subjects. Processing accounts provide evidence that the child is operating under cognitive processing limitations, but do not provide much insight into why subjects should be particularly vulnerable to omission in these circumstances. Although Greenfield and Smith's notion of informativeness remains rather vague, their pragmatic account is a first step towards offering some answers to this question. Allen (2000) attempts to provide a more theoretically adequate notion of informativeness by breaking

⁸ Although Greenfield and Smith refer to the work of Grice, they do not seem to be using his maxim of informativeness.

⁹ Greenfield and Smith stress that this notion of uncertainty is assessed in relation to the child producing the utterance and not from the point of view of the listener, whilst acknowledging that the two perspectives will often converge on the same elements.

it down into various features, and it is to her analysis of child null subjects that I now turn.

2.4.3 Informativeness 2: Allen (2000)

2.4.3.1 Overview

Allen (2000) starts from much the same basic hypothesis as Greenfield and Smith, proposing that children will ‘tend to omit arguments when the referent of the argument is maximally clear from the discourse and situational context’ (2000, p. 486). However, she attempts to work out the hypothesis in more detail by analysing the notion of informativeness in terms of what she calls ‘informativeness features’. These are features that have been claimed elsewhere (Clancy, 1993; 1997) to have some effect on the representation or omission of arguments. Allen uses these features in analysing child data from the Eskimo-Aleut language Inuktitut, a language which allows ‘rampant argument omission’¹⁰. She therefore starts by raising the question of what makes children produce any overt arguments at all. Her answer is that children will ‘produce as overt only those arguments that represent informative referents’ (2000, p. 86).

Informativeness, as Allen describes it, is a property of the referents of arguments¹¹. She tests eight informativeness features, divided into three groups: knowledge features, confusion features and search-space features. Each feature is binary, and the ‘informative value’ is the one which makes the identity of the referent less certain. For example, if the referent has not previously been talked about in the current discourse, it will be classed as ‘new’. A ‘new’ referent requires the speaker to be more informative, and the specification for the binary feature ‘newness’ will be positive in this case. Allen works from the hypothesis that the more ‘informative’ values a given referent scores, the more informative the speaker will have to be to pick it out successfully, and therefore the more likely it is to be explicitly articulated. Conversely, the more ‘uninformative’ values a referent has, the more likely it is to be omitted.

Allen’s experimental work suggests that five of the eight features that she tests have a ‘significant effect’ (p. 515) on the likelihood that an argument will be omitted. She goes on to test various possible ‘thresholds’ for the number of features that would

¹⁰ Allen claims that Inuit children omit 81.6% of arguments.

¹¹ It is unclear whether ‘informativeness’ is meant to be a property of the actual entity referred to or of the hearer’s mental representation of that referent. Some features (ABSENCE, DIFF. IN CONTEXT) are clearly properties of the referent in the physical context, whilst others (NEWNESS, CONTRAST) have to do with the ‘mental knowledge’ or ‘mental context’ of the hearer (pp. 487-488).

have to be negatively specified in order to successfully predict that an argument will be overtly realised.

2.4.3.2 The features

Allen classifies three of her informative features as ‘knowledge features’. Each relates in some way to the ‘joint knowledge of the speaker and hearer’ (2000, p. 487). The ABSENCE feature has to do with whether the referent is physically present in the context of the exchange. If a referent is physically absent, the value for this feature will be [+ABSENT], and the referent accordingly scores an ‘informative’ value for this criterion. NEWNESS is a second knowledge feature. A referent is considered ‘new’, if it has not been mentioned in the last twenty utterances, and it is then assigned the value [+NEW]. All first and second person referents are automatically assigned the value [-NEW]. The final knowledge feature, QUERY, distinguishes referents which are the ‘subject of or response to a question’ (p.488) from those which are not. A referent which is not queried, [-QUERY], requires the speaker to be less informative than one which is under question. The idea here is that [+QUERY] entities are either unidentified or newly identified, and so the hearer has ‘little mental knowledge’ (p. 488) of them. Allen’s experimental data from child Inuktitut indicates that whilst NEWNESS and ABSENCE have a significant effect on the likelihood that an argument will be omitted, QUERY does not.

As their name suggests, the three ‘confusion features’ relate to the possible confusion that may arise when more than one potential referent is present (physically or linguistically). The CONTRAST feature is assigned a positive value when the speaker is making an explicit contrast with other potential referents (this is usually indicated via gesture or tone of voice). If there is more than one potential referent in the physical context of the utterance, the DIFFERENTIATION IN CONTEXT feature will be positively specified. Similarly if there is more than one potential referent in the preceding discourse (here defined as the last five utterances) then DIFFERENTIATION IN DISCOURSE is assigned a positive value. For both the DIFFERENTIATION features, identification of potential referents may be facilitated by factors such as the semantics of the verb, markers of gender, person, number and so on. In each case, competition for the position of understood referent introduces uncertainty, and therefore an informative value is scored. Allen’s initial investigation of the confusion features indicates a significant effect on explicit argument representation for CONTRAST and DIFFERENTIATION IN CONTEXT, but not for DIFFERENTIATION IN DISCOURSE. Further investigation reveals that

DIFFERENTIATION IN DISCOURSE may indeed have an effect on the form of argument chosen, but that this is confounded by the effects of the other two confusion features.

The final category of informativeness features deals with factors relating to the size of the search space involved. Here, Allen identifies the features THIRD PERSON and INANIMACY. She claims that in typical (child) discourse, there are far fewer animate objects than inanimate objects. As a result, if the referent is inanimate, the search space will be much larger, and so an 'informative' value is assigned for this feature. Allen's analysis reveals that INANIMACY does not have a significant effect on the form of referring expression used. In fact, her definition of this feature seems to make it no more than a special case of DIFFERENTIATION IN CONTEXT, since there are likely to be more inanimate than animate potential referents.

According to Allen's data analysis, the final feature, THIRD PERSON, has the most dramatic influence on the likelihood that an argument will be overtly realised. This feature has a positive value for any third person referent, while first and second person referents receive a negative score. However, Allen acknowledges that although her results suggest that this criterion has a significant influence on the form of the argument, in Inuktitut it is generally considered ungrammatical to represent a first or second person referent overtly. This finding may therefore be at least partly accounted for by the grammar rather than by considerations of informativeness alone.

As well as considering the informativeness features separately, Allen looks at their collective effect, and carries out various analyses to investigate the predictive power of different combinations of features. She concludes that whilst the informativeness of an argument affects the likelihood that it will be expressed, and that a 'model with informativeness features as predictors' is an improvement over one without, the predictive ability of her model is far from comprehensive. As she acknowledges, 'not all informative arguments are represented by overt forms...and some uninformative arguments are represented by overt forms' (p. 512). She suggests various possible ways of accounting for this lack of consistency, including an appeal to extra-linguistic factors, possible hierarchical or cumulative effects and practical factors such as speaker mistakes, repetition or imitation.

2.4.3.3 Informativeness and English null subjects

Allen suggests that the conclusions she draws from her analysis of Inuktitut might carry over to children acquiring a non-null subject language such as English. Shaeffer (2005) takes these conclusions one step further and applies the informativeness features not

only to nulls in child English but also to the adult nulls found in the diary-style texts. Schaeffer's hypothesis is that if more than half of the features have a negative, 'uninformative' value, then the subject may be dropped¹². She puts the difference between child null subjects and those found in diary-style contexts down to the application of what she terms the 'Concept of Non-Shared Assumptions (CNSA)' (2005, p.106). The idea behind this proposal is that adult speakers are aware that their own assumptions may differ from those of their hearers, and they take this difference into account when constructing their utterances. Children in the null subject phase, by contrast, are not so consistent in recognising such differences. If they wrongly assume the hearer shares their assumptions, they may represent a referent as identifiable when it is not. As a result, children 'allow null subjects in a larger number of pragmatic contexts than adults do' (2005, p. 110).

2.4.3.4 Problems and limitations of the informativeness account

Each of Allen's features is defined in terms of binary values which are seen as corresponding directly with the properties of referents. Each value is classed as either 'informative' or 'uninformative'. I suggest that this is a substantial weakness in Allen's approach. On her analysis, a referent is treated as absent or present, new or old, differentiated or not, and these binary distinctions are used to predict the null or overt forms of the associated referring expressions. In my view, an approach based on a continuum of cases and features would be more appropriate, and would more accurately reflect what goes on in speech production. In practice, we find not only null and overt forms of referring expressions, but amongst the overt forms, we find a wide range of more or less informative forms among which the speaker must choose. Replacing binary distinctions with a continuum or gradient of 'informativeness' would make it possible to incorporate the full range of referring expressions into the account. It would also make it possible to make comparisons among different forms of referring expressions relative to the context and to each other. Consider the *NEWNESS* feature. With a binary distinction, two potential referents in a given context may end up being classified as non-new. However, one of these non-new entities may be less 'new' than the other. For example, a child and its carer have been playing with a toy bear for half an hour, when a toy dog is introduced into the play. According to Allen, the bear is a non-new entity, and as soon as the dog has been referred to once, it too becomes 'non-

¹² Subject to certain syntactic constraints of the type discussed in chapter 1: namely, 'a null subject cannot appear after a proposed constituent, or in an embedded clause' (Schaeffer 2005, p.105).

new'. As long as both have been mentioned in the last 20 utterances, then both will score the same 'newness' value. Allen's treatment does not make it possible to distinguish the bear's status as an established entity in comparison with the fairly novel dog. On her analysis, the dog's newness is an absolute value, and is in no way relative to other entities in the context.

Work by Wittek and Tomasello (2005) reveals potential problems with the use of binary distinctions in relation to the QUERY feature. They investigated which referring expressions, including nulls, were produced by German speaking children when answering different types of questions, comparing specific, general and contrast questions, as illustrated in (19)-(21) below.

(19) What happened to the broom? (specific)

(20) What do we need to get? (general)

(21) Did the clown use the vacuum cleaner?

No, the broom. (contrast)

Their results revealed that when replying to a specific question, such as (19), the children overwhelmingly used pronouns or nulls, whilst for general and contrast questions, such as (20) or (21), the answers were 'almost exclusively' (2005, p.555) lexical nouns or noun phrases. These results suggest that the form of the question significantly affects the likelihood of a null being produced, and this is not reflected in Allen's binary 'query' feature approach. Allen's analysis may provide some degree of predictive capability, but it is very narrow in its coverage. Ideally, an account of null subject usage would be integrated within a more general theory of the use of referring expressions, and this would require more than simple binary distinctions.

Apart from the limitations of using binary features, there appears to be significant overlap between the various features proposed¹³. For example, the feature THIRD PERSON is problematic in that it does not doubly dissociate from each of the other features. Whilst it is possible that an argument which is [+THIRD PERSON] may be either new or non-new, absent or non-absent, the same is not the case for the negative

¹³ This may be because Allen derives her features from a wide range of previous work by several different authors.

specification of the feature. Any argument that is [-THIRD PERSON] will necessarily also be [-NEW], according to Allen's definition of the NEWNESS feature. Thus, whenever a referent is 1st or 2nd person, it will score two 'uninformative' marks for what is, in effect, the same quality. The features are not independent.

Similar problems arise with the other 'search-space' feature. Allen's justification for treating INANIMACY and THIRD PERSON as separate features relies on the respective numbers of animate versus non-animate entities and 1st and 2nd person versus non-1st and 2nd person entities. This in turn relies on the assumption that a child is aware of the relative proportions of such entities present in the environment and wider context. Although I would not wish to dispute these assumptions outright, the more such assumptions have to be made, the more vulnerable the theory becomes. This particular assumption also requires the child to be sensitive to the hearer's knowledge state in respect of the entities to which she is referring. Witteck and Tomasello (2005) provide experimental evidence that at age 2.0 years, children are 'much less sensitive to the knowledge states of their interlocutor' than is envisaged in this assumption (2005, p. 552). If so, then it becomes difficult to argue that the child takes the hearer's search space fully into account when formulating her utterance. Even assuming that Allen is correct, there is an overlap between the INANIMACY feature and the DIFFERENTIATION IN CONTEXT feature. Indeed, as I mentioned in passing above, INANIMACY seems to be no more than a specific and predictable instance of DIFFERENTIATION IN CONTEXT. Both features are defined by the need to distinguish between more than one potential referent in the context. A referent that scores a positive value for the INANIMACY feature, should, by Allen's own reasoning, also score a positive value for the DIFFERENTIATION IN CONTEXT feature. According to Allen, inanimates require the speaker to be more informative because there are usually more of them to choose between than inanimate objects. She claims that 'if the listener knows nothing about the referent except whether it is animate or inanimate, the identity of an inanimate referent will be much less certain than that of an animate referent' (2000, p. 489). As a result, there is more potential for confusion, since there are more potential referents. This feature covers much the same ground as DIFFERENTIATION IN CONTEXT. The INANIMACY feature simply identifies one group of entities which will be likely to have an informative value for the already identified differentiation features. Including both features in a theory of argument representation risks treating the same contextual element as contributing to two separate informativeness values. Once again, the features are not independent.

Furthermore, Allen gives no indication of why these two ‘search-space’ feature distinctions are chosen over other potential dichotomies. For example, it could be argued that in a given situation there are likely to be more non-human than human entities, and yet ‘human’ is not presented or tested as a possible informativeness feature.

In sum, when considered carefully, it becomes apparent that the eight informativeness features overlap somewhat in their remits. This overlap stems partly from the lack of a clear definition of context. Allen’s notion of context is unclear and inconsistent, with the inconsistencies evident even within one feature family. For instance, the two differentiation-based members of the confusion family operate over very specific contextual domains. For DIFFERENTIATION IN CONTEXT, what is involved is the ‘immediate physical context’, even more specifically defined as the ‘portion of the room where the child is directing his/her eye gaze’ (p. 488). However, when defining the CONTRAST criterion, one feature is used to cover anything influenced by ‘potential referents in the discourse or in the shared physical or mental context’ (p. 488). No suggestion is made about why an individual feature is required for each aspect of the context in the DIFFERENTIATION cases, whilst the CONTRAST feature refers to context in general. Moreover, the definition of context given for the problematic search-space features does away with the immediate context altogether, and is instead based on levels of usage in ‘child discourse’. So it seems that Allen’s notion of context fluctuates with the various features, and lacks any constant or concrete definition.

2.4.4 A move towards relevance

The informativeness analyses put forward by Allen and Shaeffer seem to have a certain degree of descriptive value and, as Allen argues, as predictive models they are a significant improvement on preceding accounts. The fact that omissions are optional even when licit may help to explain the lower than expected level of predictive success in the informativeness model. It is noticeable that Allen’s model was much more successful at predicting which arguments would be null than with predicting which arguments would be overt. It is possible that some of these overt arguments could legitimately have been left unarticulated, but that for some reason the speaker chose not to omit them. This optionality factor (in the cases we are considering, it is never ungrammatical to replace a null with an overt form – although the converse is not the case) is missing from Allen’s account and is not directly addressed by Shaeffer. However, in her final discussion of the data, Allen hints at what I feel could be an important step in moving the work forward. She acknowledges that a discourse-

pragmatic account such as her informativeness approach could be ‘complementary to processing accounts’ (p. 515).

The competence-based approaches discussed above have nothing to say about the other omitted elements of child language, and they are unable to account for the gradual change in subject realisation as the child develops. As noted above, their reliance on the notion of a mis-set parameter is also conceptually problematic. The processing accounts, by contrast, offer some insight into why omissions occur, but are rather vague when it comes to explaining why subjects (and other elements) should be particularly vulnerable. The discourse-pragmatic analyses discussed in this section offer a richer analysis of some of the factors affecting omission, but fail to explain why omissions should occur in the first place. In the next section, I give a brief overview of relevance theory, and argue that it offers a theoretically well-developed framework in which processing and discourse-pragmatic factors can be combined to produce a more comprehensive and finer-grained account.

2.5 Relevance Theory

2.5.1 An overview of relevance theory

Relevance theory (Sperber & Wilson, 1986/95; Blakemore, 1992; Carston, 2002a; Wilson & Sperber, 2004) offers a cognitive approach to utterance interpretation based on two main principles. The first, or cognitive, principle of relevance deals with cognition in general and states that human cognition tends to be geared towards the maximisation of relevance. The second, or communicative, principle of relevance deals specifically with communication, and is formulated as in (22).

(22) Communicative principle of relevance:

Every act of ostensive communication communicates a presumption of its own optimal relevance (Sperber & Wilson, 1986/95, p. 260).

Utterances, as acts of ostensive communication, fall under this principle. The heart of the relevance-theoretic approach to pragmatics thus lies in the definitions of relevance and optimal relevance.

Sperber and Wilson define relevance as dependent on two separate factors: cognitive effects and processing effort. At any one time, an individual will hold a set of

existing assumptions about the world, from which a subset will be drawn to act as context for the processing of utterances or other inputs to cognitive processes. Cognitive effects result when an input interacts with these contextual assumptions in one of three ways: it may (a) strengthen an existing assumption, (b) contradict and eliminate an existing assumption or (c) combine with an existing assumption to yield contextual implications (conclusions derivable from input and context together, but from neither alone). To illustrate, consider the following examples. Susan is a tennis enthusiast. The week before the Wimbledon tournament begins, she decides the following:

(23) If the sun is shining on Monday I will go to the tournament.

Susan waits in anticipation and wakes up early on Monday morning. As she looks out of the window she sees that:

(24) The sun is shining.

The new information in (24) combines with Susan's existing assumption in (23) to yield the conclusion or contextual implication in (25).

(25) I will go to the tournament.

Other things being equal, the more contextual implications are derivable from a given input to cognitive processes, the more relevant that input will be.

As the tournament progresses, Susan is lucky enough to be given a ticket to the men's final. She believes that the tennis player Murray is on good form and so she holds the assumption in (26) with a fair degree of confidence.

(26) I will see Murray play in the final.

A few days into the tournament, Susan opens the newspaper and sees the headline in (26).

(27) Murray breezes past opponents in first two rounds.

The new information in (27) strengthens her assumption in (26), and achieves relevance thereby. However, two days later she sees the headline in (28).

(28) Murray knocked out in straight sets.

In this case, the new information contradicts Susan's assumption in (26) and removes her confidence in it entirely. Other things being equal, the more assumptions are strengthened or eliminated by processing a new input, the more relevant it will be. Thus we see how new information can interact with existing assumptions to result in the three types of cognitive effects, and achieve relevance thereby. Other things being equal, the greater the cognitive effects achieved, the more relevant the input will be.

The other factor in Sperber and Wilson's definition of relevance is processing effort. When cognitive effects are derived by processing an input, in a context of existing assumptions, some mental effort is expended. Other things being equal, the less effort required to derive a given set of cognitive effects, the greater the relevance of the input. Thus, when Peter asks Mary what time the train leaves, it would be more relevant for her to answer with (29) than with (30).

(29) At 5 o'clock.

(30) At 5 o'clock and Paris is the capital of France.

In processing the longer utterance in (30), Peter will use more effort, and if the information that Paris is the capital of France contributes no further effects, the overall relevance of Mary's utterance will be reduced. The level of effort required to process an utterance may be affected by a range of linguistic and non-linguistic factors, including amount of linguistic material, logical complexity of the utterance, frequency or recency of use of linguistic constructions or conceptual representations, and the size and accessibility of the context. More generally, other things being equal, the less mental effort a hearer is required to expend when processing an input to cognitive processes, the more relevant that input will be. Thus, cognitive effects and processing effort combine to produce a level of relevance for each potential input that is comparable with the effects and efforts required for other potential inputs. According to the cognitive principle of relevance, the human cognitive system tends to maximise relevance by automatically allocating attention and processing resources to those inputs

with the greatest expected relevance (Sperber & Wilson, 1986/95, pp. 118-132; Wilson & Sperber, 2004; Carston, 2002a, pp. 44-47).

However, according to Sperber and Wilson, the addressee is not entitled to expect that the speaker, in producing an utterance or ostensive stimulus, has aimed at maximal relevance, and that the utterance will therefore provide the maximum effects for the minimum effort, as compared with other possible utterances. Rather, as stated in the communicative principle of relevance, the hearer is entitled to presume that any utterance addressed to him will be optimally relevant, and to interpret it accordingly. This presumption of optimal relevance is itself part of what is communicated by an utterance, and Sperber and Wilson (1986/95, p. 270) define it as follows:

(31) Presumption of optimal relevance (revised)

- a. The ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it.
- b. The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.

This presumption in turn motivates the following comprehension procedure which, according to Sperber and Wilson (2002), is automatically used in interpreting an utterance (or other ostensive stimulus).

(32) The Relevance Theoretic Comprehension Procedure:

- a. Follow a path of least effort in deriving cognitive effects: test interpretive hypothesis (reference assignments, disambiguations, implicatures etc) in order of accessibility.
- b. Stop when your expectations of relevance are satisfied.

A consequence of the procedure is that the first accessible interpretation which makes the utterance relevant in the expected way is the one that the hearer should select. According to Sperber and Wilson (2002), the relevance-theoretic comprehension procedure is the central component of a comprehension module which automatically

constructs an interpretation for any utterance (or other ostensive stimulus) addressed to one.¹⁴

2.5.2 Relevance and child nulls

I have argued that neither existing competence accounts nor existing processing or pragmatics accounts can fully deal with the phenomenon of null subjects in child language. Each seems to tell some of the story, but none provides an adequate analysis of the full range of data. In this section, I will argue that we gain more insight into the issue if we break it down into two separate but related questions: first, why does the child's speech frequently involve omissions, and second, how does the child 'select' the items to omit? Processing accounts tend to focus on the first of these questions, whilst pragmatic accounts attempt to answer the second. I will argue that by adopting the framework of relevance theory, we will be in a position to answer both questions. In later chapters, I will apply the same principles to the adult null subjects in diary-style registers.

According to the communicative principle of relevance, when a child in the null subject phase produces an utterance, she communicates, as part of her meaning, that her utterance is optimally relevant. The hearer is therefore entitled to assume that it is at least relevant enough to be worth processing, and that it has been formulated – to the extent that this is compatible with the speaker's abilities and preferences – so as to produce as many cognitive effects as possible, for the lowest possible cost in processing effort.

This definition of optimal relevance suggests answers to both of the questions raised above. In particular, it allows us to integrate into a single account the processing constraints which encourage the child to omit some surface element or other and the pragmatic constraints which help to determine which element she will omit.

Relevance theory provides at least two points at which to incorporate processing considerations into an account of the data. In the first place, processing constraints on production are covered by the reference to the speaker's 'abilities' in the second clause of the definition of optimal relevance. In the second place, the role of processing effort in comprehension is covered by the claim that to be optimally relevant, an utterance must be at least relevant enough to be worth the hearer's processing effort, and that it has been formulated so as to yield the greatest effects, for the smallest effort, compatible

¹⁴ See also Blakemore, 1992, 2002; Carston, 2002a; Wilson & Sperber, 2002; Sperber & Wilson, 2008; Sperber & Wilson, 2008; Wharton, 2009 for fuller discussion of the relevance-theoretic approach to pragmatics.

with the speaker's abilities and preferences. I will start by considering the role of processing constraints on production.

The child's linguistic abilities and performance skills are not yet fully developed, and the child producing an utterance must try to communicate within these limitations. As we have seen, Paul Bloom and Lois Bloom provide evidence for some kind of cognitive or processing limitation in the child speaker, as a result of which sentence length or structural complexity may trigger an overload of the processing systems, leading to some surface constituent(s) giving way and being omitted. Olson's (1973) work on memory span and cognitive development suggests a further possible factor in the child's need to reduce surface form.

On this approach, the child omits surface linguistic elements to simplify the utterance and bring it within the range of her productive abilities. The extent to which subjects are dropped varies across children and from situation to situation, and relevance theory suggests an explanation for this. Just like adults, children are affected in their ability to express themselves by the state of their emotions, their physical condition and the circumstances in which they are speaking. They are also undergoing the process of acquiring a language, learning other new things about the world and developing many other skills as they grow. Relevance theory does not entail any particular account of the child's productive processing abilities, or choose between the competing accounts outlined above, but it does suggest two points that may be worth bearing in mind in developing a fuller account.

First, it predicts that the child subject-drop phenomenon makes things easier for the speaker rather than the hearer. From the point of view of an adult hearer, it makes the utterance less stylistically acceptable, and must be condoned on the grounds that the child is unable to do better, rather than seen as making a positive contribution to overall relevance. Second, the different productive processing constraints discussed in performance accounts may lend themselves to comparative treatment. Some surface forms may require more productive processing effort than others, and we may be able to assess how much overall effort would be saved by the omission of individual constituents. A speaker who is forced to omit some surface linguistic element because of productive processing constraints should omit enough elements to bring the utterance within the range of her productive abilities, while minimising the risk of misunderstanding to the extent that she is able. In considering which elements will minimise the risk of misunderstanding, the hearer's processing effort needs to be taken

into account, and I will now turn to this and the second question: how does the child 'select' the items to omit?

The presumption of optimal relevance again suggests an answer. A child aiming at optimal relevance, but forced to omit some surface element, should omit those elements whose omission is least likely to detract from overall relevance. Of course, this suggestion raises many more questions. What is it that makes the omission of some surface elements detract less from overall relevance than others? Why are subjects and certain other categories consistently omitted, whilst other elements are retained? According to the relevance theoretic account, the general answer must be that a surface element can be omitted if its content is easily inferred (causing minimal additional processing effort) at minimal risk to overall understanding. Typically, such elements will be 'given' rather than 'new', 'topic' rather than 'focus', 'theme' rather than 'rheme'. Thus, a relevance-theoretic account should be able to incorporate the results of discourse-pragmatic accounts which appeal to these notions.

As we have seen, the existing data seem to lend themselves to treatments which deal with tendencies rather than absolutes. Rule-based approaches face the problem that there are exceptions to each of their proposed rules. However, relevance theory allows us to dispense with many such rules and instead consider the data from the perspective of effort versus effect. As well as causing the speaker a certain amount of productive processing effort, an utterance will demand a certain amount of processing effort from the hearer. In return for this, it should yield an adequate range of cognitive effects. Ideally, a speaker who is forced to omit some element of the utterance because of productive processing constraints should omit those elements which (a) bring the utterance within the range of her productive abilities and (b) allow the hearer to infer her intended meaning with a minimal expenditure of extra processing effort and minimal risk to overall understanding. Thus, a full account of the child subject-drop phenomenon will need to consider both speaker's and hearer's processing effort. On this approach, it is therefore necessary to consider what factors contribute to both speaker's and hearer's processing effort, as well as to hearer's cognitive effects. It is at this point that I return to the findings from the existing accounts.

Existing accounts suggest that a number of factors contribute to both effects and effort. We have seen how negation, unfamiliar vocabulary, sentential, VP and subject length, metrical complexity and rightward complexity in an utterance increase the speaker's processing effort and appear to make omissions more likely. However, a speaker aiming at optimal relevance is most unlikely to omit a negation marker, because

its content would be extremely hard for the hearer to infer, and the risk of misunderstanding would be correspondingly great. From the hearer's perspective, the most easily dispensable surface elements would be those he is expecting to find anyway, which would therefore be particularly easy to infer. Factors such as a lack of stress, the rate of previous mention, given versus new information and topic status, which are appealed to in existing discourse-pragmatic accounts, are associated with constituents whose content is easily inferable. The more such factors are present, the more vulnerable to omission the elements are likely to be. Thus, relevance theory provides a general framework in which both processing and pragmatic factors can be analysed in terms of their contribution to overall relevance.

2.5.3 Relevance instead of Informativeness

Bearing in mind this relevance theory approach to the child null subject data, I want to return briefly to Allen's experimental work on child Inuktitut (2000). Her analysis of the data suggests that at least five of her eight informativeness features (NEW, ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT AND THIRD PERSON) have a significant influence on the form of an argument. However, I have argued that Allen's feature definitions are problematic in several ways. I now want to suggest that we can explain Allen's effects without running into the same problems, by reinterpreting her features in relevance theory terms.

A subject (or other sentential element) is vulnerable to omission when its content is easily inferred. According to the relevance theoretic comprehension procedure, hearers test interpretations in order of accessibility. We might therefore expect subjects to be more vulnerable to omission when the intended referent is the most accessible one in the discourse context (or is judged by the speaker to be so), and has no potential competitors.¹⁵ When this is the case, the hearer will test this interpretation first, and because of the lack of competitors, the risk of misunderstanding will be minimised. With this in mind, I suggest that Allen's informativeness features should be viewed as a (non-exhaustive) list of factors which are likely to correlate with the accessibility of a referent in a given discourse context.¹⁶ Recall that on Allen's approach, a negative specification for a feature contributes to the argument's overall informativeness and therefore makes it more likely to be overtly expressed. This follows from Greenfield

¹⁵ I leave to one side for the moment the issue of egocentrism and the possibility that the child assesses accessibility from her own perspective rather than from the perspective of the hearer.

¹⁶ In section 5.3.2 I discuss some further factors which have also been shown to contribute to the accessibility of a referent.

and Smith's pragmatic principle: 'omit uninformative arguments'. How, then, might these features be reinterpreted in relevance theory terms?

According to Allen's analysis, two of her knowledge features have a significant effect on whether or not an argument is explicitly articulated: NEW and ABSENCE. Assuming that hearers test interpretations in order of accessibility¹⁷, we would expect non-new and non-absent referents to be more accessible than potential referents which are new or absent. Thus, we would predict that null subjects are more likely when the referent is non-new and non-absent. An advantage of this approach is that potential referents are assessed for accessibility relative to one another and to the discourse context, rather than against a binary criterion defined in somewhat arbitrary terms (e.g. by counting preceding utterances), as in Allen's account. Given two candidate referents which would both count as NEW according to Allen's feature definition, relevance theory allows us to treat one as more NEW, and therefore perhaps less accessible, than the other¹⁸. This relative approach to accessibility is appropriate if we take seriously the idea that the hearer will test interpretations in order of accessibility. How accessible a potential referent is at a given point will fluctuate as the discourse progresses and the linguistic and physical context change. A similar story can be told for the 'confusion' features, including both of the DIFFERENTIATION features. If there are competing potential referents in either the linguistic or the physical context, then the speaker will need to use a more explicit form of referring expression in order to distinguish between them, and thus minimise the risk of misunderstanding. If two potential referents are competitors, omitting the referential expression entirely is likely to lead to confusion and therefore risk putting the hearer to gratuitous effort.

The final significant confusion feature is CONTRAST. By definition, an entity that scores positively for the feature CONTRAST is not going to be the single most accessible potential referent in the context and so, once again, the risk of misunderstanding is increased. As Allen points out, the fact that there is a contrast is often explicitly signalled via tone of voice or gesture, and an explicit signal of this sort will have to be realised on an overt subject. In these cases, although the intended referent is in contrast with a competing candidate referent, it is explicitly indicated by the speaker, and so becomes highly accessible for the hearer.

¹⁷ In this chapter I will remain fairly vague about what it means to be 'accessible'. This issue will be discussed in more detail in the following chapters.

¹⁸ In certain discourse contexts, a newly introduced entity is highly accessible due to its novelty. Thus, we are not so concerned with defining entities as 'new' or 'non-new' but rather with their salience in the discourse context.

Finally, consider Allen's 'search-space' features. As discussed above, INANIMACY was not found to be significant and could in any case be reinterpreted as a special case of DIFFERENTIAL IN CONTEXT. By contrast, THIRD PERSON was shown to have a significant effect. I have argued that Allen's definition of this feature seems to overlap with that of the feature NEW. However, even if we disregard this problem, we might expect first and second person referents (i.e. the child and her interlocutor) to be highly accessible during a discourse. Both are, by definition, actively involved in the discourse and should therefore be highly salient in the discourse context. If this is the case, then they will be highly accessible and the speaker is more likely to be able to pick them out successfully by use of a null form.

Thus, there is a sense in which all of Allen's factors might be seen as 'search-space' factors. A hearer is testing interpretations in order of accessibility, and the speaker is constructing her utterance with this in mind. Each of the informativeness features singles out a factor that is likely to affect the accessibility of a referent. It is not the features themselves that are significant, but the effect they are likely to have on the overall accessibility of the intended referent. The higher the 'informative' score for an intended referent, the more unlikely it is that it will be the most accessible entity in the discourse context. The information carried by the subject may provide important clues to the speaker's intended referent, and so leaving the subject null is more likely to affect overall relevance. As a result, the chance that the subject will be fully articulated is, therefore, increased.

To sum up, we might view Allen's informativeness features as indicating the degree to which a certain referent is likely to be accessible in a given context. When it is not highly accessible, the speaker will, if possible, tend to avoid the use of nulls in order to provide the information necessary to constrain the search space so that the most accessible referent is the intended one. This is not to say that Allen's list of informative features is in any way exhaustive (I discuss further factors in 5.3.2) but merely that we can explain her experimental results within the relevance theory framework appealing to the notion of accessibility.

2.5.4 Concluding remarks

In this chapter, I have considered various approaches to the phenomenon of child null subjects in non-*pro*-drop languages. Neither the competence, processing or pragmatic accounts surveyed can fully explain the data. However, I propose that an analysis which appeals to both processing limitations and pragmatic factors can provide us with

a more comprehensive and explanatory account, and I suggest that adopting the framework of relevance theory allows us to do just that.

During the null subject stage the child is operating under processing limitations which have an impact on her abilities. As a result, she produces an utterance which is the most relevant one compatible with her abilities, and the elements which are articulated will be those least likely to affect the overall relevance of the utterance. Thus, the null subject phenomenon is a result of the speaker aiming for optimal relevance while constrained by processing limitations, and as such should fit into a relevance theoretic account of the use of referring expressions in general. I have argued that the accessibility of the intended referent in the discourse context interacts with the abilities of the speaker (most significantly in this case, the processing limitations of the child speaker) to make certain elements of an utterance vulnerable to omission. In the next chapter, I consider how we might characterise what it means to be accessible, and how this might fit into a more general theory of the use of referring expressions. In chapters 4 and 5, I develop my approach to referring expressions, and in chapter 7, I return to the issue of child null subjects in the light of my relevance theory account of null subjects in chapter 6.

Chapter 3: Referring and Accessibility

3.1 Introduction

In the previous chapter I argued that in child language, processing constraints can result in the omission of certain elements from the surface form of the utterance, and that the items which are most vulnerable to omission are those which contribute least to the overall relevance of the utterance. Since relevance is a function of cognitive effects and processing effort, the accessibility to the hearer of the intended referent (and hence of the intended interpretation) is a crucial factor for the speaker constructing an utterance. From this perspective, null subjects are simply one in a range of possible means by which a speaker can achieve reference. With this in mind, I will devote this chapter to some existing pragmatically-orientated accounts of referring expressions. In particular, I consider two accounts which focus on the relation between the form of a referring expression and the accessibility or givenness of the intended referent, by Mira Ariel (1990; 1994; 2001), and by Jeanette Gundel and her colleagues (Gundel, Hedberg, & Zacharski, 1993; Gundel & Mulkern, 1998; Gundel, 2003).

3.2 Contextual accounts

Existing pragmatically-orientated analyses of referring expressions investigate links between the form of a referring expression used and the nature of the context from which the hearer is to identify the referent. Clark and Marshall (1981) treat context as involving ‘mutual knowledge’ between speaker and hearer. Mutual knowledge, as defined by Schiffer (1972), is an infinitely regressive form of shared knowledge in which speaker and hearer must not only share certain assumptions, but know that they share them, and know that they know that they share them, and so on ad infinitum¹. According to Clark and Marshall, mutual knowledge of a certain assumption may be established via three routes: membership in a community which shares this assumption, co-presence in a physical environment which provides evidence for it, or co-presence in a linguistic environment which provides evidence for it, and the context for utterance comprehension is drawn from such knowledge. Clark and Marshall’s central claim is

¹ For a critique of this notion, and an alternative account, see Sperber & Wilson (1986/95, chapter 1)

that each of these types of mutual knowledge is associated with a certain type of definite reference:

Deixis corresponds to physical copresence; anaphora corresponds to linguistic copresence; and proper names correspond to community membership (Clark & Marshall, 1981, p. 42).

Mira Ariel classifies the associated forms of referring expressions as ‘Linguistic Givenness Markers, Physical Givenness Markers, and Knowledge Givenness Markers’ respectively (Ariel, 1990, p. 6, see also Ariel 1985).

Approaches of this type rely on what Ariel terms ‘geographic context-form correlations’, where the form of referring expression used is directly tied to the context type to which it points. Whilst such accounts seem to make correct predictions in many cases, Ariel (1990) argues that ‘we cannot establish a one-to-one correspondence between form and function according to the context-type’ (1990, p. 7). To illustrate, she discusses the ‘stringent requirements’ that Clark and Marshall (1981, p. 44) treat as applying to the use of pronouns which in their view require the linguistic co-presence of the intended referent. Ariel (1990, p. 8) provides examples showing that, while this may often be the case, pronouns may, in fact, relate to any of the three context types:

(1) a. Encyclopaedic Knowledge:

Sherlock Holmes to Watson: The butler did *it* (the murder, eliminating incriminating evidence, etc)

b. Physical Context (drinks have been poured out).

Sherlock Holmes to Watson: The butler did *it* (the pouring of the drinks).

c. Linguistic Context:

What a cruel murder. The butler did *it* (the murder, not the eliminating of incriminating evidence).

In (1c) the pronoun ‘it’ is interpreted as co-referential with the ‘cruel murder’ already mentioned in the linguistic context. In (1a) this linguistic context is

missing and the pronoun is identified via the hearer's encyclopaedic knowledge. The referent of the pronoun in (1b), uttered after drinks have been poured, is identified via the physical context and is taken to be the act of pouring drinks. The fact that this range of interpretations is possible suggests that analyses based purely on 'geographic' context-form correlations will be unable to account for the full range of data. Ariel concludes that 'any theory that attempts to simply correlate between linguistic markers of Givenness and context-types cannot be maintained' (1990, p. 10).

Taking this position as her starting point, Ariel goes on to develop an account of referring expressions which she calls Accessibility theory². Accessibility theory moves away from the idea that choice of referring expression is driven by context type, and instead proposes that referring forms encode information about the Accessibility of the intended referent, which is therefore not just a matter of pragmatics, but 'forms part of the grammar of natural languages' (1990, p. 100). In the next section I provide an overview of Ariel's theory, and outline some associated implications and problems.

3.3 Accessibility theory

3.3.1 Overview of Accessibility theory

Accessibility theory (Ariel, 1985; 1988; 1990; 1991; 1994; 2001) provides an account of referential choice and interpretation in which the degree of Accessibility of the mental representation of the intended referent is marked in the grammar. In this view, all referring expressions are seen as broadly anaphoric: as Ariel puts it, Accessibility theory does 'away with the referential-anaphoric distinction' (1990, p. 7). When a speaker uses a referring expression, she is signalling to her hearer how accessible she believes the mental representation of the intended referent (the antecedent) to be for that hearer in that particular context. Referring expressions are therefore viewed as markers which function by 'cueing the addressee on how to retrieve the appropriate mental representation in terms of degree of mental accessibility' (Ariel 2001, p. 31).

According to Ariel, Accessibility is the 'determining principle which accounts for the choice of referential forms' (1988, p. 69); she defines it as a complex 'graded psychological notion' (1994, p. 37) which results from the interaction of several different factors. Properties of both the mental representation of the referent itself and

² Following Ariel (1990), I will capitalize 'Accessibility' when referring to Ariel's theory-specific notion, as outlined in section 3.3.1.

the relationship between the anaphor and the antecedent contribute to the overall degree of Accessibility, and therefore to the choice of referring expression (Ariel 2001). Broadly speaking, these properties fall into four categories: saliency (i.e. whether the referent is salient or not, 'mainly whether it is a topic or non-topic'), competition (i.e. 'the relative saliency of an entity when compared with other potential referents'), distance (i.e. the distance between the antecedent and the anaphor) and unity (i.e. whether the antecedent is within 'the same frame/world/point of view/segment or paragraph as the anaphor') (Ariel, 1990, pp. 28-29). Of these, the first two relate to the representation of the intended referent, and the latter two to the relationship between the referent and its antecedent. On this account, the status of the antecedent as a global topic, local topic or non-topic will affect its saliency, and hence its Accessibility, as will the number of representations competing for the role of antecedent. Similarly, degree of Accessibility will be affected by the distance between the anaphor and its antecedent, and by any coherence relations that may exist between the two.

Memory plays a crucial role in determining how Accessible a mental representation will be, and Ariel adopts a Parallel Distributed Processing (PDP) model of memory (McClelland, 1994). This model allows for 'indefinitely many distinctions in degree of activation' (Ariel 1990, p.15), which in turn allows indefinitely many degrees of Accessibility. Degree of activation in memory is a crucial factor in how easily the intended mental representations will be retrieved. According to Accessibility theory, referring expressions act as 'price tags' (1990, p.16) indicating to the hearer how much processing effort he should expect to expend on retrieving the intended representation.

3.3.2 A scale of Accessibility

Accessibility theory treats referring expressions as 'guidelines for retrievals' (1988, p.68), which signal to the hearer how Accessible he should expect the intended referent to be. Although Ariel moves away from a geographic notion of context, she formulates her theory in relation to the same three context types: general knowledge, physical environment and linguistic material. Whereas in previous approaches referring expression types were seen as directly associated with one of the three context types, Ariel treats referring expressions as encoding different levels of Accessibility, which are in turn related to different context types. Information in the immediately preceding linguistic discourse has the highest level of Accessibility, and will be quickly and easily available to the hearer. Information drawn from the physical environment is associated

with mid-level Accessibility, and finally, general knowledge shared between the speaker and hearer is associated with low levels of Accessibility. Thus, the context types form a hierarchy, and the association between context type and referring expression proposed by the ‘geographic’ accounts is seen as ‘an adequate description of *unmarked* usage in *initial* retrievals’ (1990, p.69). However, by introducing the concept of Accessibility, Ariel claims she can also account for marked and non-initial uses. As shown by the examples in (1) above, the link between context type and referring expression does not necessarily hold in all cases. Still, Ariel claims that ‘the very same graduation of marking is maintained’ in all cases (1990, p.71). The relative degree of Accessibility signalled by the different referring forms stays constant across uses, and from this Ariel derives her Accessibility marking scale. It should be noted that although Ariel often talks of referring expressions as markers of low, mid or high accessibility, this is an acknowledged simplification of a much more complex and fine-grained Accessibility scale.

Full name + modifier	Low Accessibility
Full (‘namy’) name	↓
Long definite description	
Short definite description	↓
Last name	
First name	↓
Distal demonstrative + modifier	
Proximal demonstrative + modifier	↓
Distal demonstrative (+NP)	
Proximal demonstrative (+NP)	↓
Stressed pronoun +_ gesture	
Stressed pronoun	↓
Unstressed pronoun	
Cliticized pronoun	↓
Extremely High Accessibility Markers	High Accessibility

Table 1: Accessibility Marking Scale (English) (Ariel 1990, p.73)

Ariel's formulation of Accessibility is intended as both universal and language specific. Whilst the claim that referring expressions are hierarchically ordered markers of relative Accessibility is considered a universal, each language is seen as making the association between Accessibility and its particular inventory of referring expressions in its own way. The Accessibility scale for English is given in Table 1. Thus, according to Accessibility theory, when speakers of English use an unstressed pronoun, they are signalling that the intended antecedent should be highly Accessible to the hearer. If, however, they choose to use a long definite description, they would be signalling that the hearer should be retrieving a referent with relatively low Accessibility.

The fact that the scale is both universal and also language specific is seen as falling out naturally from Ariel's next important claim regarding Accessibility. According to the theory, while the degree of Accessibility is conventionally encoded as part of the meaning of the referring expression, the correlation between Accessibility and the associated form is not arbitrary. Rather, it is driven by the interaction of three factors: the informativity of the referring expression (i.e. the amount of lexical information it carries), its rigidity (i.e. how close it is to 'pointing to one entity unequivocally in a potentially ambiguous context') and its attenuation (i.e. its phonological size). Ariel sees informativity as the most important of these factors, and claims that the 'more informative, the more rigid and the least attenuated the form, the lower Accessibility it marks' (1991, p. 449). In her view, the combination of these three criteria can largely motivate the scale in table 1. Stressed and unstressed pronouns, for example, are distinguished by degree of attenuation. Although a stressed pronoun encodes the same semantic information as an unstressed pronoun, it is a less attenuated form, and is therefore associated with lower Accessibility. Whilst the interaction of the three factors is designed to demonstrate how the relation between form and Accessibility is not entirely arbitrary, Ariel acknowledges that for some of the distinctions, a certain amount of arbitrariness is inevitable. For example, the difference between the distal demonstrative 'that' and the proximal demonstrative 'this' cannot be accounted for via Ariel's three factors alone. Both terms are equally rigid, equally attenuated and, according to Ariel, equally informative. She therefore treats the relative positions of 'this' and 'that' on the English scale as reflecting an arbitrary, language-specific distinction.³

³ In the next chapter I will discuss how this observation forms an important part of Ariel's justification for Accessibility theory, and how relevance theory might offer a different perspective on the data.

3.3.3 Accessibility theory and relevance theory

Ariel positions Accessibility theory firmly within the relevance theory pragmatic framework, and claims that neither Accessibility theory nor relevance theory alone can fully account for the use and interpretation of referring expressions. Accessibility theory is presented as ‘a very helpful facilitating device’ (1990, p. 182) which functions as a ‘useful tool serving Relevance assessments’ (1990, p. 86). The principles of relevance are needed to ‘select among equally accessible potential referring expressions’ (2001, p. 38), and so Accessibility theory on its own is insufficient. According to Ariel, the converse is also true, and relevance theory ‘cannot account for the distribution of referring expressions without the mediation of Accessibility theory’ (1990, p. 84). She briefly acknowledges the possibility that her three defining criteria (informativity, attenuation and rigidity) could potentially be ‘replaced by a Relevance theory account, possibly rendering Accessibility theory redundant’ (1990, p. 82). However, she argues that certain pairs of referring expression forms are ‘indistinguishable as to the amount of information they impart’, and that as a consequence ‘the accounts given by Relevance and Minimization⁴ are at a loss to explain why the more costly version is ever used’ (1990, p. 85).

According to Ariel, relevance theory, with its twin factors of effort and effects, is unable to distinguish between definite descriptions and their corresponding demonstratives, as in (2)-(3), between distal and proximal demonstratives, as in (3)-(4), or between stressed and unstressed pronouns, as in (5)-(6).

(2) *The plane* flew everyday from Miami to New York.

(3) *This plane* flew everyday from Miami to New York.

(4) *That plane* flew everyday from Miami to New York.

(5) Jane_i kissed Mary_j and then she_i kissed Harry.

(6) Jane_i kissed Mary_j and then SHE_j kissed Harry.

⁴ The Minimization principle (Levinson, 1998) is a general ‘least effort’ strategy which drives the speaker to produce the shortest possible form of an utterance.

Accessibility theory, on the other hand, can motivate the differing interpretations by claiming that ‘conventionalised Accessibility markings exist’. Discussing this proposal, Reboul (1997), suggests that Ariel’s argument rests entirely on the premise that, in each of these pairs, both forms convey exactly the same information (apart from Accessibility information). If we accept this premise, then (in Reboul’s view) Ariel’s conclusion would indeed follow, and a strong case would be made for some form of Accessibility marking.

However, Reboul points out that by accepting this premise, we are committing ourselves to the view that, for example, (7) and (8) convey exactly the same information.

(7) This cat is hungry.

(8) That cat is hungry.

For Ariel’s argument to hold, the only difference in the information conveyed by (7) and (8) must be about the Accessibility of the intended antecedent. But as Reboul points out, this does not seem to be the case: the demonstrative determiners encode information about ‘the relative distance between the speaker and the object designated’, and she argues that this spatial information ‘cannot be equated with Accessibility’ (Reboul 1997, p.11).⁵ If this is the case, then, contrary to Ariel’s claim, the forms in (7) and (8) can be distinguished in terms of the information they convey, and relevance theory may have something to say about their distribution. Although this argument does not entirely rule out the possibility that referring expressions are conventionalised markers of Accessibility in the way that Ariel suggests, if we can account for their distribution without appealing to the extra machinery of an Accessibility scale, then such an account should be preferred. In the next chapter, I reconsider Ariel’s arguments against the possibility of a relevance-only account of referring expressions, and argue that by treating referring expressions as encoding procedural meaning that is not itself analysable in terms of Accessibility, we can do away with Accessibility theory entirely.

⁵ In the next chapter I look more closely at how this information might be characterised.

3.3.4 Problems and limitations

Apart from questioning the assumptions on which Ariel builds her argument for the introduction of Accessibility theory, Reboul (1997) discusses some further problems with the Accessibility approach.

As noted above, Ariel treats all referring expressions as broadly anaphoric, claiming that ‘in all cases an addressee looks for antecedents which are themselves mental representations’ (1990, p. 6). As a result she overlooks the fact that they ‘can indeed refer and do so fairly frequently’ (Reboul 1997, p. 10). This leaves her with little to say about cases where, for example, demonstratives are used to refer to something in the physical context which the hearer has not yet noticed, and it also raises questions about how her account might deal with indefinites. Reboul concludes that Ariel’s notion of Accessibility is ‘simple and monolithic’ (1997, p.17), and she develops this point by looking more closely at the data which Ariel uses to support her theory. Although Ariel describes Accessibility as a function of four factors (saliency, competition, distance and unity), her empirical analysis is based on only one of these: distance between antecedent and anaphor in a text. Furthermore, her examples are drawn from a corpus that contains only single-speaker, written sources, such as newspaper articles and short stories. Reboul questions whether Ariel would have got similar results if she had included transcripts of spoken discourse in her sample. She concludes that ultimately what is missing from Ariel’s account of referring expressions is reference itself.

Finally, both Ariel and Reboul seem to have overlooked a major factor which distinguishes different types of referring expression, and which is available to relevance theory: the amount of effort required to process these expressions. On the assumption that stressed pronouns require more effort to process than unstressed pronouns (Sperber & Wilson, 1986/95, pp. 202-217; Wilson & Wharton, 2006), we have a means to distinguish utterances (5) and (6) (‘Jane kissed Mary and then she/SHE kissed Harry’). Similarly, the higher frequency of use of ‘the’ compared with ‘this’ or ‘that’⁶ should mean that it requires less effort to process, and so provides a means to distinguish between examples (2) and (3) (‘*The/this plane* flew everyday from Miami to New York’). Processing effort will play a crucial role in my relevance-based account of these examples, and also in my analysis of diary-style null subjects.

⁶ A search of the British National Corpus returns 6055159 instances of ‘the’, compared with 1119443 instances of ‘that’ and only ‘457821’ instances of ‘this’. The contrast is more striking, given that the search did not exclude non-demonstrative uses of ‘that’ (e.g. introduction of a relative clause).

3.3.5 Special uses and special problems

3.3.5.1 Special uses

Accessibility theory treats the level of Accessibility encoded by a referring expression as correlating with the Accessibility of the intended referent in a given discourse context. However, Ariel concedes that this relation is not always simple, and she acknowledges that there are examples which ‘on the face of it seem to pose counter-examples to Accessibility theory’ (1990, p.198). On some occasions, a speaker will appear to ignore the Accessibility of the intended referent, and choose a referring expression which marks a higher or lower level of Accessibility than is associated with the intended referent at that point in the discourse. According to Ariel, such uses ‘encourage an addressee to derive specific additional contextual implications’ (1990, p 199), and she defines three categories of such uses.

First, there are cases which Ariel terms ‘insincere’, where the speaker treats a ‘non-accessible entity as if it were accessible’ (Ariel, 1990, p. 199)⁷. Consider the contrast between the distal demonstrative and the definite description in (9) and (10) respectively.

(9) That holiday we spent in Cyprus was really something, wasn’t it?

(10) The holiday we spent in Cyprus was really something, wasn’t it?

According to Ariel, the sense of ‘here and now’ associated with the use of the demonstrative in (9) creates a vividness effect. The speaker ‘makes an appeal to some shared experience, attempting to bring the past to life’ (1990, p. 199).

Second, the speaker may use a referring expression which marks low accessibility when the intended referent is highly accessible in the discourse context. To illustrate, consider examples such as (11), where a definite description is used when a pronoun would have been acceptable.

(11) When John came home this afternoon, the son of a bitch broke all the windows.

⁷ It is not clear what Ariel means by ‘non-accessible’ here, or how example (9) would work at all if the intended referent were ‘non-accessible’. It is also not clear why Ariel distinguishes between cases where non-accessible entity is treated as accessible, and cases where the speaker ignores the low accessibility of an entity and refers to it using a relatively higher marker. In my view these would seem to be slightly different versions of the same strategy.

According to Ariel, such uses are 'aimed at focusing on some particular aspect of the referent' (1990, p. 201).

Finally, there are cases where the referring expression used is associated with a higher level of Accessibility than might perhaps be expected. According to Ariel, use of high accessibility markers such as pronouns can result in an 'empathy effect'. She goes on to propose that the Accessibility scale is also a scale of empathy, and that high accessibility forms imply a level of closeness or intimacy.

Alongside these cases, there are also occasions where the use of a high Accessibility marker is 'unjustified'. Ariel claims that such uses carry negative connotations. She illustrates this with a discussion of what she terms 'conventional Accessibility raising devices' (1990, p. 204). For example, a proper name may be prefaced with a demonstrative determiner, as in (12).

(12) This Henry Kissinger really is something!

In Ariel's view, this has the effect of 'cancelling the familiarity assumption associated with names' (1990, p. 205). When this is unjustified - in other words, when the referent cannot reasonably be assumed to be unfamiliar - a negative, derogatory effect is generated.

3.3.5.2 Special problems

Ariel's discussion of the special uses of referring expressions is on the whole rather vague. However, she touches on something that I will explore further in the next chapter: the idea that referring expressions may contribute more to what is communicated by an utterance than just a referent. I will consider examples of the type discussed by Ariel, and argue that they do indeed contribute to the derivation of implicatures. However, Ariel's treatment of these examples raises some problematic issues about her proposal to analyse referring expressions as Accessibility markers, and about the role they play in reference resolution.

The special effects outlined by Ariel are the result of a departure from the normal correspondence between the Accessibility marker and the representation of the intended referent. It follows, therefore, that, in these cases, the hearer must make a crucial decision when processing the referring expression. In some cases, the intended

interpretation will be derived if the hearer proceeds normally, and takes the expressions as providing ‘no more than guidelines for retrievals’ (Ariel 1988, p.68). In others, he must override these guidelines, and retrieve some alternative effect instead. It is unclear how the hearer decides which strategy to follow. Although Ariel refers to acceptable and unacceptable uses of Accessibility markers, she gives only one hint as to when the speaker might judge a use unacceptable: this is when ‘the local context make(s) it manifest that the unusual referring expression to some extent follows from it’ (1990, p 201). However, she gives no further details on how the context would guide the hearer in this way, and it remains unclear how the hearer is to judge that a referring expression is to be treated as unusual in some cases, and when he should take it as purely and simply a signpost to the intended referent.

Once the hearer has decided that extra effects are intended, he may still face potential problems. According to Ariel, use of a higher than expected Accessibility marker may lead to empathy effects, or it may produce negative connotations. Ariel claims that the interpretation depends on whether the use is ‘unjustified’ or not. However, she gives little indication of how the hearer should judge a use as unjustifiable, except to say that it happens ‘when a speaker cannot plausibly be assumed to be unfamiliar with the referent (1990, p. 207). So, whilst on the one hand Ariel is arguing for a treatment of referring expressions as ‘guidelines for retrievals’, on the other hand, her account relies on the hearer already knowing how accessible the intended referent is: otherwise, he will be unable to decide on the correct interpretation of the expression. But if the hearer needs to know how accessible the referent is in order to interpret the referring expression, then the level of accessibility does not need to be signalled.

In the next chapter, I will argue that these problems can be overcome by developing an analysis of referring expressions which is fully integrated into the relevance theory pragmatic framework. On this account, rather than having to decide whether a particular use is appropriate or justified, the hearer simply follows the relevance theoretic comprehension procedure, and tests interpretations in order of accessibility until his expectations of relevance are met.

3.4 Gundel and colleagues: The Givenness hierarchy

3.4.1 Overview

Like Ariel, Gundel, Hedberg and Zacharski (1993), (henceforth GHZ), attempt to explain why speakers choose certain referring expressions in certain discourse contexts, and how hearers identify the intended referent from those expressions. Whereas Ariel claims that levels of Accessibility are encoded as part of the conventional meaning of referring expressions, GHZ base their approach on the cognitive status (involving ‘information about location in memory and attention state’ (1993, p. 274)) of the intended referent. They propose that different forms of referring expressions ‘signal different cognitive statuses...thereby enabling the addressee to restrict the set of potential referents’ (1993, pp. 274-5), and they identify six cognitive statuses which form a Givenness hierarchy, reproduced here in Table 2.

In focus	Activated	Familiar	Uniquely identifiable	Referential	Type identifiable
>	>	>	>	>	
{it}	{ this that this N	{that N}	{the N}	{indefinite this N}	{a N}

Table 2: The Givenness Hierarchy (Gundel et al 1993, p. 275)

Each cognitive status is associated with a different form or forms of referring expression, and the table shows the necessary and sufficient conditions for the appropriate use of those forms. For a referent to satisfy the condition of being ‘type identifiable’ the hearer need only be able to access a representation of the type of object involved. An intended referent will count as ‘referential’ if the hearer is able to retrieve an existing representation of that referent, or construct a new representation of it. If the hearer is able to ‘identify the speaker’s intended referent on the basis of the nominal alone’, then it is said to be ‘uniquely identifiable’. Continuing up the scale, if the intended referent is represented in the hearer’s memory then it is said to be ‘familiar’, and if it is ‘represented in current short-term memory’, then it is ‘activated’. The ‘activated’ status is necessary and sufficient for the use of the bare demonstratives ‘this’ and ‘that’ and the proximal complex demonstrative ‘this N’. Finally, an intended referent which is not only in current short-term memory, but is also the current centre of attention is said to ‘in focus’ (1993, p. 276-230).

On the surface, this approach seems very similar to Ariel’s. According to GHZ, however, it is the nature of the relationship between the various statuses on the

hierarchy that makes their approach different. These cognitive statuses are in a one-way implicational relationship, so that satisfaction of the conditions for a particular status entails satisfaction of the conditions for all the statuses further down the scale. Thus, if an intended referent is ‘activated’, then it is by definition also ‘familiar’, ‘uniquely identifiable’, ‘referential’ and ‘type identifiable’. By contrast, satisfying the conditions for being ‘activated’ does not entail that the referent also satisfies the conditions for being ‘in focus’, and so on.

It follows from this approach that a form associated with a relatively low cognitive status may be used to refer to an intended referent which satisfies the conditions for a higher status. Consider (13), taken from Gundel et al (1993, p. 296).

- (13) Dr Smith told me that exercise helps. Since I heard it from A DOCTOR, I’m inclined to believe it.

At the point where the phrase ‘A DOCTOR’ is processed, the doctor in question is represented in current short-term memory, and is arguably the current centre of attention. So, at the very least, the conditions for the ‘activated’ status are met, and it might also be claimed that the referent is ‘in focus’. However, according to GHZ, use of the indefinite article is perfectly acceptable in this example, even though it is associated in the Givenness hierarchy with ‘type identifiable entities’. This is because, as noted above, the speaker intending to refer to an ‘activated’ referent may choose a form associated with ‘activated’ or with any status below ‘activated’ on the scale. By comparison, consider the conditions that would need to be satisfied for (14) to be uttered appropriately.

- (14) That dog next door kept me awake last night.

The use of the term ‘that dog’ indicates that the referent is at least ‘familiar’. According to GHZ, this means that the addressee has a representation of that particular dog in long-term memory. If (14) were to be uttered in a discourse context where the hearer does not know that the speaker’s neighbours have a dog, then GHZ’s account predicts that the utterance will, at the very least, seem odd. In such a discourse context, the referent would fail to satisfy the necessary conditions for a ‘familiar’ entity, and replacing ‘that dog’ with the referring form ‘the dog’ from one step down the scale feels much more appropriate.

The Givenness hierarchy, like Accessibility theory, is claimed to be both universal and language specific. According to a survey carried out by GHZ, cross-linguistic application of the approach reveals that the six identified statuses are sufficient, although not always necessary, for describing referring expression use in the languages included in the study. English was the only language in the group which required the status 'referential', and Japanese, Russian and Chinese were found only to require four of the remaining five statuses. This cross-linguistic survey also revealed that the correlation between linguistic form and cognitive status was not arbitrary: the forms at the top of the scale, and therefore associated with 'in focus' entities, tended to have the least phonetic content in all the languages studied.

In sum, the Givenness hierarchy makes predictions about the 'highest' form of referring expression that may be used to pick out a particular entity in a given discourse context. In many cases, this will leave the speaker with several choices of referring expression. Theoretically, whenever an entity is 'in focus', the speaker has the full range of referring expressions to choose from, since conditions for all these statuses will automatically have been satisfied. Conversely, when a hearer processes an indefinite description, he could, in theory, take it to apply to anything that satisfies the nominal part of the description:

'a dog'...is appropriate only if the addressee can be assumed to know the meaning of the word 'dog' and can therefore understand what type of thing the phrase 'a dog' describes (1993, p. 276).

Despite this apparent flexibility, GHZ argue that the choice between possible referring expressions is not random. They propose that an interaction between the Givenness hierarchy and Grice's conversational maxims (Grice, 1975; 1989) can explain why any particular form is selected on any given occasion. In this way, they claim to be able to fully account for the distribution of referring expressions in language use.

3.4.2 The Givenness hierarchy and Grice

As noted above, the implicational nature of the Givenness hierarchy means that there will often in principle be several choices of referring expression available to the speaker. If an entity is in focus, it can be picked out by using any of the referring expression types. Similarly, any potential referent could be picked out by use of an indefinite article. Sitting at the bottom of the scale, and requiring only that the hearer can identify what type of thing is being referred to, indefinite articles should be

available in all instances. However, GHZ's survey of the distribution of referring expression forms confirms that the choice between the various available forms is not random, and that certain clear patterns emerge. For example, although theoretically available in all instances, indefinites tend to occur only when the referent is, at most, 'referential'. Similarly, whilst 'in focus' entities may be referred to using any type of expression, discourse analysis reveals that they are usually picked out using forms associated with the more restrictive cognitive statuses. GHZ argue that these patterns result from interaction between the Givenness hierarchy and Grice's maxims of quantity, given here in (15).

- (15) a. Make your contribution as informative as required.
- b. Do not make your contribution more informative than is required. (Grice, 1989, p. 26).

Since each cognitive status on the Givenness hierarchy entails the statuses below it, as we move up the hierarchy the cognitive statuses become more restrictive, and there will be fewer and fewer entities in a discourse context that satisfy the necessary conditions. Thus, the referring expressions associated with statuses higher up the scale can be said to be more informative, since they narrow the set of potential referents further. It follows from Grice's first quantity maxim (Q1), given in (15a), that when linguistic terms form a scale of this sort, use of a particular term generally implicates that conditions for the use of higher terms are not met.

Scalar implicatures of this sort arise and have been much discussed in the Gricean pragmatic literature (Levinson, 1983; 2000, Horn, 1984; Carston, 1997; Noveck & Sperber, 2007; Geurts, 2010). Examples (16) and (17) illustrate the implicature standardly associated with the <some, all> scale.

(16) Some of my friends are softballers.

(17) All of my friends are softballers.

Utterance (17) is more informative than (16). According to Q1, if the speaker is in a position to use the more informative version, then she should do so. As a result, if a speaker utters the less informative (16), she implicates that (as far as she knows) (17)

does not pertain. The hearer is then entitled to infer that not all of the speaker's friends are softballers. Crucially, this is not a logical entailment of (16), since it can be cancelled without contradiction, as in (18).

(18) Some of my friends are softballers, in fact, all of them are!

According to GHZ, similar effects arise with the Givenness hierarchy, and some of the distributional facts described above can be explained by appeal to Grice's first quantity maxim. Thus, use of an indefinite article typically implicates that the referent is not uniquely identifiable. Gundel and Mulken (1998, p. 25) discuss the following example (19), which I have abbreviated for reasons of space.

(19) From Minneapolis to Mankato I was sitting next to a black woman.

According to the Givenness hierarchy, the indefinite article in (19) signals that the referent is at least type identifiable. The speaker should therefore expect at least to be able to access a representation of the type of entity under discussion: in this case 'a black woman'. Furthermore, via Q1, (19) conversationally implicates that the intended referent is not uniquely identifiable. The hearer should therefore not expect to be able to assign a unique representation to the woman.

The statistical correlation found in GHZ's survey between 'in focus' referents and use of stressed pronouns and zeros is also predicted by the interaction between the Givenness scale and Q1. If an entity is in focus, then the most informative way to refer to it is by using a form which restricts the set of potential referents accordingly. This leads GHZ to argue that instances of what they call focus shift can be explained via the Givenness hierarchy. Consider the example in (20) (Gundel, Hedberg, & Zacharski, 1993, p. 298).

(20) ...going on back from the kitchen then is a little hallway leading to a window, and across from the kitchen is a big walk-through closet. On the other side of *that* is another little hallway...

At the point when the demonstrative 'that' is uttered, both 'the kitchen' and the 'closet' are activated. However, only 'the kitchen' is in focus. According to GHZ, 'demonstratives not only don't require the referent to be in focus, but often implicate

that the referent is not currently in focus' (1993, p. 299). As a result, the hearer is guided to a referent which is activated but not in focus. If we replace 'that' with 'it', then 'the kitchen' is the only possible referent, since it is the only 'in focus' entity.

However, GHZ also point out that use of a particular form associated with a certain cognitive status doesn't always implicate that conditions for higher statuses are not met. They demonstrate this by comparing definite descriptions with complex demonstratives. Not only does use of a definite determiner not implicate that a demonstrative determiner would not have been appropriate, but when conditions for both statuses are met, the lower form is used in the majority of cases. GHZ argue that this is due to the influence of second quantity maxim (Q2) ('Do not make your contribution more information than is required'). Definite noun phrases - unlike pronouns - contain descriptive material, which, according to GHZ, is often enough by itself to identify the intended referent. As a result, explicit signalling of a more restrictive cognitive status would be superfluous, and a speaker complying with Q2 should avoid it. They go on to claim that 'since most references which are uniquely identifiable in a discourse are also at least familiar, explicitly signalling a status higher than uniquely identifiable would be more informative than required' (1993, p. 300).

3.4.3 Limitations of a Gricean analysis

Whilst interaction between the Givenness Hierarchy and Grice's maxims of quantity may go some way towards explaining the selection and interpretation of referring expressions, I want to suggest that this approach is limited in its explanatory power, and overlooks certain important properties of referring expressions and the contribution they make to overall utterance interpretation.

Reconsider (13) ('...since I heard it from a doctor, I'm inclined to believe it'). GHZ appeal to this example to illustrate their claim that use of a relatively low form does not necessarily implicate that conditions for higher forms are not met. Whereas use of the expression 'a black woman' in (19) is seen as implicating that the speaker does not know the woman in question⁸, non-identifiability is not part of the encoded meaning of an indefinite descriptions. Thus, for instance, the indefinite noun phrase 'a doctor' in (13) does not entail that the doctor in question is not uniquely identifiable, familiar and so on. According to GHZ, since what is relevant in this utterance is the

⁸ I will go on to argue that in such examples lack of knowledge is only one possible explanation for the speaker's under-informativeness. It may be the case that the speaker knows, but does not wish to disclose, further information, or it may be that she simply does not judge the extra information as likely to make a positive contribution to the relevance of the overall utterance.

property of being a doctor, rather than the identity of the doctor, an indefinite description is acceptable. I believe that with this observation they are touching on an important point. Consider the alternative utterance in (21).

(21) Dr Smith told me that exercise helps. Since I heard it from him, I'm inclined to believe it.

It might be argued that 'a doctor' in (13) and 'him' in (21) are used to pick out the same person, namely Dr Smith. However, the two utterances clearly communicate very different overall meanings. As GHZ point out, in (13), the speaker suggests that she believes Dr Smith because he is a doctor. In (21), however, she suggests that she believes him because of the particular person he is. Since (21) is acceptable, we know that the referent is 'in focus', and hence, according to the entailment relations of the Givenness scale, any referring expression type lower on the scale should also be acceptable. However, whilst both versions are acceptable, they convey different overall meanings. The Givenness scale gives us no way of explaining why this should be. According to the GHZ analysis, utterance (13) will not carry Q1 implicatures because it is the property of being a doctor rather than the identity of the doctor that is relevant. I suggest that the fact that the hearer is being encouraged to consider the property of being a doctor as relevant is itself an important part of what is communicated by (13).

According to relevance theory, overall utterance interpretation involves constructing an appropriate hypothesis about the contextual assumptions the hearer is expected to use. In order to understand the clause 'Since I heard it from a doctor' in (13) as explaining why the speaker believes that exercise helps, the hearer must construct a hypothesis about the intended contextual assumptions. For utterances (13) and (21), the beliefs attributed to the speaker, and used as contextual assumption will be along the lines of (22) and (23), respectively.

(22) All doctors give good advice.

(23) Dr Smith gives good advice.

Since assumption (23) is a special case of (22), utterance (13) provides an explanation for the speaker's decision to believe Dr Smith, whilst also communicating that she trusts doctors in general. However, (23) does not entail (22). Although it too provides an

adequate explanation, it would be perfectly possible for the speaker to trust Dr Smith, but to have no opinion about the trustworthiness of doctors in general. Hence, (13) and (21) are qualitatively different in what they communicate. GHZ overlook this point. They fail to explain how the speaker of (13) conveys (22), whilst relying on it to explain the absence of a Q1 implicature in this case.

As this example illustrates, GHZ claim that, whilst the Givenness hierarchy may interact with Grice's Q1 and Q2 maxims to yield quantity implicatures, such implicatures do not necessarily arise in all cases. In their view, Q2 implicatures frequently arise when definite descriptions are used to refer to an entity that is 'familiar'. Although definite descriptions signal that the entity need only be 'uniquely identifiable', GHZ's analysis relies on the assumption that 'most references which are uniquely identifiable in a discourse are also at least familiar' (1993, p. 300). Using the 'familiar' form, 'that N', would therefore be more informative than is required. However, if the difference between 'uniquely identifiable' and 'familiar' so often fails to be exploited, then there appears to be little justification for positing two different levels on the hierarchy in the first place. Indeed, GHZ's definitions of these two levels are similar: in both cases the intended referent may be identified via a representation of it in the hearer's memory. In addition, 'uniquely identifiable' entities may be identified by the 'descriptive content...encoded in the nominal itself' (Gundel, Hedberg, & Zacharski, 1993, p. 277). It follows from the Givenness hierarchy that when something is familiar, it is also uniquely identifiable. According to GHZ, the nature of uniquely identifiable entities means that they are also usually familiar. Thus, the two statuses overlap considerably and it becomes unclear how or why the two levels should be distinguished at all.

Furthermore, if, as GHZ claim, Q1 implicatures do not arise automatically in all cases, and Q2 implicatures may arise when certain forms are used in certain contexts, then we need some explanation for how a hearer judges when it is appropriate to derive an implicature and when it is not. It seems that by appealing to Grice's maxims, the Givenness account tacitly acknowledges that its own treatment of the cognitive processes underlying the speaker's choice of referring expression does not go far enough.

Apart from these problems the Givenness/Gricean account is also fairly limited in its explanatory power. The only inferences allowed under the GHZ (1993) account are those relating to cognitive statuses, and therefore to reference resolution. I will argue in the next chapter that the speaker's choice of referring expression can and does result in

a wide range of other inferential effects, and that these cannot be explained by appeal to cognitive status alone.

3.4.4 The Givenness hierarchy and relevance theory

As noted above, the GHZ framework as originally presented in 1993 relies on interaction between the Givenness hierarchy and Grice's maxims of quantity. I have argued that this account leaves many questions unanswered about how hearers recognise which implicatures to derive and when. In later work, Gundel and Mulhern (1998) demonstrate that the GHZ account also falls short of explaining the full range of data and judgements. For example, according to the Givenness hierarchy, the use of the definite article 'the' signals that the referent should be at least uniquely identifiable. In discourse, however, the definite article tends to be used for entities that are represented in short term memory, and are therefore at least 'activated'. Consider (24), from Gundel and Mulhern (1998, p. 29).

(24) I closed the bedroom door. Later, I noticed that the door was open.

In (24), the definite description 'the door' is most naturally interpreted as referring to the previously mentioned bedroom door. Whilst this door is indeed uniquely identifiable, it is only one of many doors that are likely to be uniquely identifiable to the hearer. The GHZ framework has no obvious way of explaining why the previously mentioned door should be selected in preference to any other uniquely identifiable door. What is needed, they conclude, is some explanation of how the intended interpretation is chosen from all those with the appropriate cognitive status. Here, they turn to relevance theory and find that it offers an explanation of how this might take place:

...in processing an utterance an addressee will assign the most easily accessible interpretation that provides adequate contextual effects. This theory would therefore predict that in determining the intended interpretation of a referring expression, the addressee would choose the interpretation from the most accessible context, as long as this is consistent with the meaning directly encoded in the form itself (Gundel & Mulhern 1998, p.32).

An explanation along these lines might seem to open up the possibility that relevance theory alone can explain the use and interpretation of referring expressions without

appeal to the notions used in the Givenness hierarchy, and in later chapters, I will argue that it does. However, like Ariel, Gundel and Mulkern argue that relevance theory on its own is inadequate to deal with the full range of data, since in their view, it has ‘nothing to say about differences in the way particular forms (e.g., ‘a’, ‘the’, ‘this’, ‘that’, ‘it’) constrain possible interpretations of referring expressions’ (1998, p. 33).

Consider the referring expressions in (25) and (26), and the associated judgements.

(25) A restudy of pareiasaurs reveals that these primitive reptiles are the nearest relatives of turtles.

(26) ??A restudy of pareiasaurs reveals that the primitive reptiles are the nearest relatives of turtles.

Gundel and Mulkern (1998) and Gundel (2003) discuss these examples, and argue that the Givenness hierarchy explains the data in a way not available to a relevance only approach. In (26), the definite description ‘the primitive reptiles’ is judged unacceptable on a reading where ‘primitive reptiles’ is coreferential with ‘pareiasaurs’, whilst the corresponding complex demonstrative in (25) is judged acceptable. What makes (26) unacceptable in their view is that the most accessible interpretation of the definite description ‘the primitive reptiles’ is a group of primitive reptiles that is not coreferential with (although it may include) the pareiasaurs. In (25) and (26), the definite description and complex demonstrative forms are seen as carrying the same semantic information which, according to Gundel, ‘constrains possible interpretations to primitive reptiles, but...provides no information about which primitive reptiles are intended’ (Gundel 2003, p. 6). The key to understanding the differences in acceptability and interpretation must therefore lie with the determiner. According to the Givenness hierarchy, the demonstrative ‘these’ signals that the referent is at least ‘activated’ in the hearer’s cognitive environment. Gundel and Mulkern claim that the complex demonstrative in (25) is acceptable because there is only one potential referent which satisfies the ‘activated’ criterion (namely the set of pareiasaurs).

By contrast, the definite article ‘the’ in (26) signals that the referent should be at least ‘uniquely identifiable’. This cognitive status is lower down the scale of Givenness, and is therefore less restrictive. As a result, there will be more potential referents which satisfy this requirement. According to Gundel, there might well be

other groups of primitive reptiles, including the group of all primitive reptiles, which satisfy the condition of being uniquely identifiable. The referring expression in (26) should therefore be more difficult for the hearer to resolve, and, if anything, will lead to a different interpretation.

According to the Givenness hierarchy a referring expression signals that the intended referent meets the necessary and sufficient conditions for a particular cognitive status. This restricts potential referents to just those that meet these conditions. Gundel and Mulken (1998) then introduce the idea that considerations of relevance drive the choice between these referents. According to this modified analysis, neither the Givenness hierarchy nor relevance theory can fully account for the data without appealing to the other. In the next chapter, I will consider this claim in more detail, and argue that this is not correct. I will argue that relevance theory, supplemented by semantic information which makes no appeal to notions of either Givenness or Accessibility, can account for the full range of reference cases, and so there is no need to introduce theory-external scales or hierarchies.

3.5 Against a scalar analysis: the case of names

The Accessibility and Givenness accounts both depend on the assumption that referring expressions form a hierarchy or scale. Whether they define the points on the scale as representing a level of Accessibility or a cognitive status, they share the idea that referring expressions encode this information, and signal it to the hearer as a guide to reference resolution. In this section, I present some objections to such scalar accounts, and suggest instead that the appropriateness of a referring expression varies with the discourse context in which it is used. Discourse context, I will argue, affects the felicitous use of a referring expression much more than its place on a hierarchy or scale. The data I discuss will also raise the issue of how considerations of style might influence the use of referring expressions. I will return to this point in much more detail in chapters 4 to 7.

Ariel presents the Accessibility scale in 3.3.2 as a language-specific manifestation of universal principles, which is affected by the informativity, attenuation and rigidity of referring expressions in English. As a result of language-specific factors, the scale may vary in particular details from language to language. For example, if a language does not use zero forms, the accessibility scale will reflect this, and will differ from the scale for a language such as Italian, which does use zero forms.

In this section, I use the case of proper names in English to argue that a scalar approach is inadequate to account for the data. Leaving aside any cross-linguistic differences, the accessibility level associated with the use of names does not remain constant even within a single language. Names occupy four places on Ariel's Accessibility scale for English (Table 1). 'Full name + modifier' sits at the very bottom of the scale, closely followed by 'full ('namy') name'. Moving up the scale, and skipping over long and short definite descriptions, we find 'last name', followed by 'first name'. According to this scale, then, use of a last name signals that the referent is less accessible to the hearer than if a first name had been used. According to Ariel, 'last names are less ambiguous than first names' (1990, p. 81), and first names are 'not such good disambiguating tools. Hence they must refer to relatively highly accessible entities' (1990, p. 40). At first sight these claims seem quite plausible. Certainly, for the kind of academic works and newspapers from which Ariel draws much of her data, they are likely to hold. However, I will argue that this is not always the case, and that the appropriate form will vary not only with the accessibility of the referent, but also with the specifics of the discourse context, including social conventions about the use of proper names. By 'discourse context', I mean the set of assumptions ('thoughts treated by the individual as representations of the real world' (Sperber & Wilson, 1986/95, p. 2)) accessible in the cognitive environment of the communicator and audience. Bearing this in mind, I will conclude that Ariel's claim that there is an 'association of specific forms with specific levels of Accessibility', and that 'degree of Accessibility dictates formal choices' (1990, pp. 75-76) is too strong.

Consider the referring expressions in (27)-(30), listed in order of Accessibility from lowest to highest, as indicated on Ariel's scale.

(27) The pop singer, Kylie Minogue.

(28) Kylie Minogue.

(29) Minogue.

(30) Kylie.

Accessibility theory predicts that (27) should be used when the referent is of very low accessibility, (28) when the referent is slightly more accessible, and so on until the high

accessibility form in (30). Now, consider (31) and (32) as headlines on the cover of a magazine.

(31) Minogue has new boyfriend.

(32) Kylie has new boyfriend.

Assuming that a headline of this sort is taken as a discourse-initial utterance, and that the set of potential referents includes all the celebrities and public figures within the remit of the magazine's interest group, we might expect the lower Accessibility marker (i.e. 'Minogue' to be preferred, since, as Ariel, claims, 'last names are less ambiguous than first names'. On this account, (31) should be the preferred headline. However, this prediction is not necessarily correct, and I suggest that there are at least two reasons for this.

First, in the given discourse context (based on the world of celebrity gossip), there are two candidate referents with the last name 'Minogue', but only one with the first name 'Kylie'. In this discourse context, therefore, it is simply not the case that a last name is less 'ambiguous' than a first name. In an alternative discourse context, the situation may be reversed. If, for example, there are two or more girls with the first name Kylie in the same school class, an utterance of (32) might well fail to pick out a unique referent, and the speaker would have to use a different form. It is therefore not possible to make generalisations about the pragmatic appropriateness of a referring form based on Ariel's notion of Accessibility alone. Rather, the choice of an appropriate referring expression must be made relative to a discourse context, and is driven not only by the form of the referring expression itself, but also by properties (e.g. rarity of use of the name, number of competing potential referents) of the specific expression which vary depending on the discourse context in which it is used. This type of interaction between discourse context and the level of accessibility associated with a referring expression is not provided for within Accessibility theory.

Secondly, stylistic or social factors seem to play an important role in the choice of referring expression. Again, this is something not explicitly allowed for in Ariel's framework. In certain circumstances, for instance, it seems to be socially appropriate to pick out a referent by their last name only (e.g. in academic citation, court reporting or political commentary), whilst in others it would seem odd (e.g. in a discussion amongst friends or in a gossip column). Although in some cases the choice will be linked to the

formality of the register, this is not always so. Soccer fans discussing a game, for instance, may be likely to refer to the players by their surname only, despite the informality of the conversation.

We can see how the need to distinguish between potential referents might interact with stylistic or social conventions in examples (33) and (34).

(33) Williams is my favourite tennis player.

(34) Venus is my favourite tennis player.

Although tennis players are, by convention, most often referred to by their last names, in the case of two successful tennis-playing sisters, this strategy is likely to lead to confusion. In a discussion of sportsmen and women, the utterance in (33) is unlikely to enable the hearer to assign reference to a unique individual. However, the relative rarity of 'Venus' as a first name, specifically within a discourse context involving knowledge of famous tennis players, means that an utterance of (34) will be likely to be felicitous. What this suggests is that each use of a referring expression in discourse depends not only on the location of the referring expression on a scale or hierarchy, but on interaction between the specific referring expression used and the discourse context in which it is uttered.

By categorising referring expressions by type, and failing to consider their appropriateness relative to the discourse context in which they are used, Accessibility theory leaves itself unable to account for differing uses of different instances of the same referring expression type. It cannot explain why it is inappropriate to refer to people by their first name in some discourse contexts and by their last name in others. Just as (31) ('Minogue has new boyfriend') would be inappropriate in an informal style of publication, so (35) would be inappropriate in an academic journal, no matter how accessible the mental representation of Noam Chomsky might be.

(35) Noam now writes more often on politics than on linguistics.

Such examples suggest that treating Accessibility marking as part of the 'inherent definition' of a referring expression may be too strong. The resulting treatment does not allow for stylistic conventions, differing levels of uniqueness and rigidity, and most importantly, the interaction of referring expressions with the discourse context in which

they are used. Ariel concedes that in certain circumstances some referents may ‘possess a salient and privileged status for us’ (1990, p. 41). However, it is not clear how such exceptions would fit with the notion of a ‘universal’ scale, or what the role of the ‘scale’ would be if the scalar relationships can be so readily disregarded.

3.6 Concluding remarks

Accessibility theory and the Givenness hierarchy each attempt to give a pragmatically orientated account of the use of referring expressions in natural discourse. Both claim to integrate with the relevance theoretic pragmatic framework, whilst arguing that relevance theory alone cannot account for all of the relevant data.

Both Accessibility theory and the Givenness hierarchy characterise referring expressions as encoders of semantic information relating to the status of the intended referent in the hearer’s cognitive environment. For Ariel, referring expressions signal level of Accessibility, and for Gundel et al, they signal cognitive status. Over the next two chapters I will develop an account which moves away from the idea that referring expressions encode this sort of information. Rather, I will assume that hearers following the relevance theoretic comprehension procedure always test the most accessible potential referent in the discourse context first, and that speakers choose referring expressions with this in mind. I will present an account on which referring expressions encode procedural and/or conceptual information which restricts the set of potential referents to a point where the most accessible interpretation is the intended one. On my account, the contribution of a referring expression is not determined by the type of expression used (definite description, pronoun, distal complex demonstrative etc), but by the concepts and procedures that it encodes, and in each case the encoded content interacts with the specific discourse context to contribute to the overall interpretation of the utterance.

Ariel appears to be moving in a similar direction when, in an overview of her work on referring, she explicitly describes Accessibility theory as a ‘procedural analysis’ (2001, p. 29). However, I do not take her use of the term ‘procedural’ as signalling any development in her overall approach. She outlines the role of the procedure as being to ‘instruct the addressee to retrieve a certain piece of Given information from his memory by indicating to him how accessible this piece of information is to him at the current stage of the discourse’ (2001, p. 29). Thus, reference still depends on encoded information about the accessibility of the intended

referent. Even if we allow that Ariel's account is procedural, the procedures she is proposing differ significantly from the ones I will propose.

In this chapter, I have outlined problems and limitations associated with Ariel's and Gundel's scalar accounts. I suggest that many of these problems stem from the fact that these accounts focus almost entirely on reference resolution, and hence on the derivation of the explicature (in relevance theory terms). By focusing on the contribution of the referring expression to what is explicitly communicated, they fail to consider how referring expressions might contribute to what is implicitly communicated (i.e. the intended context or cognitive effects). They also underestimate the role played by the discourse context and stylistic or social considerations in the use and interpretation of referring expressions.

In the next chapter, I will argue not only that relevance theory can account for the data without the need for these auxiliary scales and hierarchies, but that in doing so, a much richer picture of the role of referring expressions is revealed. As well as contributing to what is explicitly communicated by an utterance, the choice of referring expression can affect what is implicitly communicated. More generally, the speaker's choice of referring expression is driven by considerations of relevance, and will contribute to the overall interpretation of the utterance in context.

Chapter 4: Referring Expressions and Relevance Theory

4.1 Introduction

4.1.1 Arguments against a relevance-only account

In chapter 3, I considered two pragmatically-orientated approaches to referring expressions which claim to be compatible with relevance theory. However, whilst endorsing the principles of relevance theory, both Ariel (1990) and Gundel (2003) argue that relevance theory alone cannot fully explain a speaker's choice of referring expression. As supporting evidence, they cite a range of data that they claim cannot be accounted for in terms of relevance alone. In the next two chapters, I will try to show how these examples can be handled in a purely relevance-based account.

As a brief reminder of the type of evidence put forward by Gundel and Mulken (1998) and Gundel (2003), consider again the examples introduced in chapter 3, and given here again as (1) and (2).

- (1) A restudy of pareiasaurs reveals that *these primitive reptiles* are the nearest relatives of turtles.
- (2) ??A restudy of pareiasaurs reveals that *the primitive reptiles* are the nearest relatives of turtles.

The definite description 'the primitive reptiles' in (2) is judged to be less acceptable than the complex demonstrative 'the primitive reptiles' in (1) when intended to pick out the set of pareiasaurs. According to Gundel (2003), it is harder to resolve reference in (2), and 'the most accessible interpretation here is one that is not coreferential with *pareiasaurs* (though it may be a set that includes pareiasaurs)' (2003, p. 128). In other examples, however, definite descriptions and complex demonstratives seem to be interchangeable. For instance, (1) is the opening line of an academic paper (Lee, 1993), and is immediately followed by (3).

- (3) *The two groups* share numerous derived characteristics.

(4) *These two groups* share numerous derived characteristics.

As Gundel and Mulhern (1998) point out, the version in (4), where the definite description ‘the two groups’ is replaced by the complex demonstrative ‘these two groups’, is not only equally acceptable, but also has the same interpretation as (3). In both cases the most accessible interpretation is that in which ‘two groups’ picks out the previously mentioned turtles and pareiasaurs. Any account of the use of referring expressions must be able to explain why definite descriptions and complex demonstrative are sometimes interchangeable, as in (3) and (4), and sometimes not, as in (1) and (2).

As discussed above, Gundel and Mulhern (1998) argue that relevance theory alone cannot account for these data, but that combining the Givenness hierarchy with relevance provides an explanation. Following the GHZ hierarchy, the definite article in (2) only restricts the set of potential referents to those which are at least ‘uniquely identifiable’. Recall that according to GHZ, ‘each status on the hierarchy is a necessary and sufficient condition for the appropriate use’ of the associated form (Gundel, Hedberg, & Zacharski, 1993, p. 275). On the intended interpretation, the definite description ‘the primitive reptiles’ refers to the previously mentioned set of pareiasaurs, but there are other sets of primitive reptiles, including the set of all primitive reptiles, which are also uniquely identifiable. As a result, the utterance in (2) does not pick out a unique referent, and the intended interpretation fails. By contrast, the complex demonstrative ‘these primitive reptiles’ in (1) requires the referent to be not only ‘uniquely identifiable’ but also ‘activated’. Since the set of pareiasaurs are the only activated potential referents in the discourse context, (1) is acceptable and comprehensible.

In contrast, the cognitive status ‘uniquely identifiable’ is sufficient to determine a unique referent in both (3) and (4). In these examples, there is only one set of ‘two groups’ which is uniquely identifiable, namely the pareiasaurs and turtles, and so the definite determiner ‘the two groups’ in (3) is acceptable. By definition, a referent that is ‘activated’ is also ‘uniquely identifiable’, and so the use of the complex demonstrative, ‘these two groups’ in (4) is also acceptable. Accordingly, Gundel and Mulhern claim that the Givenness hierarchy, with its notion of encoded cognitive statuses, explains the judgements in (1)-(4), and that without it, relevance theory cannot account for these data.

As noted above, Ariel (1990) sees a similar problem and cites similar data, again claiming that relevance theory alone cannot distinguish between certain pairs of referring expressions: definite determiners and complex demonstratives, proximal and distal demonstratives, and stressed and unstressed pronouns. In her view:

Relevance theory...cannot account for the distribution of referring expressions without the mediation of Accessibility theory (1990, p. 84).

In this chapter and the next, I will argue that, contrary to the claims of Ariel and Gundel et al, the existing relevance theory framework can fully explain the data in (1)-(4), and shed light on the differences between the pairs of expressions identified by Ariel.

My strategy will therefore be to offer an alternative relevance-based account of the data as presented and discussed in the literature. Before embarking on my analysis, however, I would like to enter a caveat: it is not clear that (1) and (2) bring out a genuine contrast between definite descriptions and complex demonstrative forms. To illustrate this point, compare (5)-(8) with the examples in (1) and (2).

- (5) A restudy of pareiasaurs reveals that *the primitive reptiles* were mainly found on the African mainland.
- (6) A restudy of pareiasaurs reveals that *these primitive reptiles* were mainly found on the African mainland.
- (7) A reexamination of the pareiasaur's bones reveals that *the primitive reptile* consumed calcium-rich food during its lifetime.
- (8) A reexamination of the pareiasaur's bones reveals that *this primitive reptile* consumed calcium-rich food during its lifetime.

Here the versions with definite descriptions seem to be just as acceptable as those with complex demonstratives, and this suggests that the contrast between (1) and (2) does not depend only on the choice of definite NP type. In fact, one might go further and argue that the classification of (2) as unacceptable on the given interpretation is too strong. The use of a definite description as an anaphoric epithet is acceptable in parallel cases, such as those in (5) and (7), and I suggest that in certain contexts, it may also be acceptable in example (2). Ultimately, I will argue that the contrast in acceptability

between each of these pairs depends on the inferences that the various forms encourage the hearer to draw via the procedural meaning they encode. However, there does seem to be some intuitive contrast in acceptability between (1) and (2), and this needs to be explained. I will therefore begin by setting aside the data in (5)-(8), accepting that there is a contrast, and considering how it might be explained in relevance theory terms. I will then return to it at the end of the chapter.

I want to argue that, far from being unable to account for the alternations and contrasts in these examples, relevance theory gives us more insight into the varied contributions that referring expressions may make to what is explicitly and implicitly communicated. The vital contribution of referring expressions to implicit content is generally overlooked by existing accounts of reference choice, and I will argue that this is what explains the judgements in (1)-(8). Crucially, I claim that the role of referring expressions is to guide the hearer, not just to an intended referent, but to an intended overall interpretation.

According to relevance theory (Wilson & Sperber, 2004, p. 615), the construction of an overall interpretation of an utterance comprises three distinct sub-tasks, which are typically performed in parallel:

- (9) Constructing an appropriate hypothesis about explicit content (EXPLICATURES) via decoding, disambiguation, reference resolution, and other pragmatic enrichment processes.
- (10) Constructing an appropriate hypothesis about the intended contextual assumptions (IMPLICATED PREMISES).
- (11) Constructing an appropriate hypothesis about the intended contextual implications (IMPLICATED CONCLUSIONS).

I will argue that referring expressions may contribute to all three of these subtasks, and that as a result they affect not only what is explicitly communicated, but also what is implicitly communicated. I will develop this argument within a relevance theoretic framework which incorporates a distinction between procedural and conceptual meaning (Blakemore, 1987; 2002; Wilson & Sperber, 1993; Wharton, 2009), and which takes seriously the claim that hearers automatically follow the relevance theoretic comprehension procedure in constructing an overall interpretation.

In defending Accessibility theory, Ariel claims that ‘Relevance accounts assume that in reference retrievals a speaker guides an addressee by making sure he picks the right antecedent based on eliminating ‘wrong’ choices of competitors’ (Ariel 1990, p. 85). I will argue that this is a misconstruction of relevance theory, and that by taking the relevance theoretic comprehension procedure seriously, we find that far from relying on the elimination of ‘wrong’ choices, the production and comprehension of referring expressions is driven by the same principles that drive communication in general. I will start with a brief reminder of some of the basic relevance theory principles and definitions introduced in 2.5 above.

4.1.2 The relevance theoretic comprehension procedure

According to the communicative principle of relevance, repeated in (12) below, any utterance addressed to someone conveys a presumption of its own optimal relevance, as defined in (13):

(12) Every ostensive stimulus conveys a presumption of its own optimal relevance

(13) An optimally relevant utterance will be:

- a. Relevant enough to be worth the audience’s processing effort.
- b. The most relevant one compatible with the communicator’s abilities and preferences.

According to the relevance-theoretic comprehension procedure in (14), the way to find an overall interpretation is to follow the path of least effort in looking for implications and other cognitive effects.

- (14) a. Follow a path of least effort in deriving cognitive effects. Test interpretive hypotheses (disambiguations, reference resolutions, implicatures etc.) in order of accessibility.
- b. Stop when you expectations of relevance are satisfied.

Having derived enough implications, at a low enough cost, to satisfy his expectations by making the utterance not only worth his processing effort, but the most relevant one

compatible with the speaker's abilities and preferences, the hearer should stop. Thus, the first interpretation that satisfies the hearer's expectations of relevance should (if all goes well) be the one the speaker intended to convey.

Although the relevance theoretic comprehension procedure is explicitly stated in the relevance literature, it should be remembered that it does not have to be learned or acquired, but follows directly from the Communicative Principle of Relevance and the presumption of optimal relevance, and is seen by Sperber and Wilson (2002) as the central component of a comprehension module which automatically constructs a hypothesis about the speaker's meaning. Any analysis that aims to offer a relevance theory account of the use of referring expressions must be compatible with the comprehension procedure. What, then, does it mean to say that interpretive hypotheses are tested 'in order of accessibility'? In the previous chapter I showed how Ariel's Accessibility theory claims that levels of accessibility are encoded by referring expression types, and how Gundel, Hedberg and Zacharski (1993) take a similar code-based approach to cognitive statuses on their Givenness hierarchy. I have argued that these approaches are problematic, and I now want to claim that they are also unnecessary. The 'most accessible' interpretation of an utterance is, by definition, simply the first that the hearer comes across on the path of least effort. If this interpretation satisfies his expectations of relevance, then according to the relevance-theoretic comprehension procedure, the hearer should stop. It follows that a speaker aiming for optimal relevance should do her best to formulate her utterance so that, if all goes well, the intended interpretation is the first interpretation the hearer will find that makes the utterance relevant in the expected way. I want to argue that the speaker's choice of referring expression is a vital component in this process, and in the rest of this chapter and in the next, I will consider what role each of these choices might play in a relevance-based analysis of utterance interpretation.

4.1.3 The conceptual-procedural distinction and the role of procedural meaning

Relevance theory takes a largely computational-representational approach to cognition and communication (Fodor, 1983; Wilson & Sperber, 1993). On this approach, communication involves the construction and manipulation of a series of representations, phonetic, phonological, syntactic and conceptual, which are linked by a series of computations. Thus, the phonetic form of an utterance is seen as undergoing a linguistic decoding process to yield a semantic or logical form (or in the case of ambiguity, a set of logical forms). As decoding proceeds, the logical form is enriched

via inferential pragmatic processes, including reference assignment, to yield a fully propositional form known as the proposition expressed or (if it is part of what the speaker is taken to communicate) the basic explicature. This basic explicature may itself be inferentially enriched to yield a series of higher-level explicatures carrying speech-act or attitudinal information. Moreover, as described above, explicatures combine with contextual assumptions to provide input to further inferential processes yielding a series of contextual implications or implicatures (Sperber & Wilson, 1986/95; Carston, 2002; 2004). Thus, utterance interpretation involves a complex interaction between (linguistic and conceptual) representations and (linguistic and inferential) computations.

A central feature of relevance theory is the claim that while most regular content words (e.g. ‘reptile’, ‘pareiasaur’) encode concepts that figure in conceptual representations, some expressions are better analysed in procedural terms, as contributing to the computational aspect of utterance interpretation. For instance, Blakemore (1987) argued that a range of non-truth-conditional discourse connectives (e.g. ‘but’, ‘so’, ‘also’) are best seen as encoding, not information which contributes directly to conceptual representations, but information about the type of inferential computations the hearer is expected to go through in constructing an overall interpretation (Blakemore, 1987). This laid the foundations for an important theoretical distinction between conceptual and procedural encoding, which has played a major role in relevance-theoretic accounts of both verbal and non-verbal communication (Blakemore, 1987; 2002; Wilson & Sperber, 1993; Wharton, 2009). On this approach, conceptual encoding yields conceptual representations that figure directly in the explicatures that provide the input to further inferential computation, while procedural encoding places constraints on the types of representations to be constructed or the computations that are to take place (Blakemore, 1987; 2002; 2007; Wharton, 2003; 2009; Hall, 2007). As Blakemore puts it, expressions that encode procedures ‘do not encode a constituent of a conceptual representation (or even indicate a concept), but guide the comprehension process so that the hearer ends up with a conceptual representation’ (2002, p. 91).

To illustrate the kind of work standardly claimed to be done by procedurally encoded meaning, consider the following example¹:

¹ See Blakemore (1987, pp. 81-91) and (2002, pp. 89-90) for further discussion of ‘after all’ and ‘so’.

(15) a. Clare is a vegetarian.

b. She doesn't like meat.

The two utterances could stand in many different inferential relations, but the route the hearer is expected to follow can be indicated by use of the non-truth-conditional discourse connectives 'so' and 'after all' as in (16) and (17).

(16) Clare is a vegetarian. So she doesn't like meat.

(17) Clare is a vegetarian. After all, she doesn't like meat.

Most linguists and philosophers who have looked at such examples agree that the truth conditions of (16) and (17) (and hence the propositions they express) are the same. Both are true if and only if Clare is a vegetarian and she doesn't like meat. However, the inferences that the hearer is encouraged to draw are constrained by the choice of connective. In (16), the suggested inference is that Clare's dislike of meat is a consequence of her vegetarianism. However, in (17), the inference is that her vegetarianism is in some way a consequence of her dislike of meat. Thus, although it does not contribute conceptual information to the proposition expressed by the utterance, the procedural meaning of the discourse connective clearly has an important influence on the inferential phase of interpretation and the implicatures the hearer is encouraged to derive.

On the basis of such examples, it might be tempting to see the conceptual / procedural distinction as lining up with the distinction between truth-conditional and non-truth conditional meaning. However, Wilson and Sperber (1993) argue that this is not the case, and that four distinct categories of meaning emerge: conceptual and truth-conditional, conceptual and non-truth-conditional, procedural and truth conditional and procedural and non-truth conditional. The first of these categories is the most easily illustrated. Most 'content' words, including nouns and verbs, can be assumed to encode conceptual meaning which contributes directly to the truth-conditional content of an utterance. Discourse connectives such as those in (16) and (17) are plausibly seen as encoding procedural meaning which does not affect the truth-conditions of an utterance. According to Wilson and Sperber (1993) and Ifantidou-Trouki (1993), sentence adverbials such as 'seriously', 'frankly' and 'confidentially' exemplify a third

possibility. Although they do not contribute to the truth-conditions of the utterances in which they occur, they are best seen as encoding conceptual information.

(18) Confidentially, I don't think Bob will get the job.

(19) Seriously, I'm pleased Bob didn't come.

Intuitively, utterance (18) is true if and only if the speaker doesn't think that Bob will get the job, and it therefore seems that 'confidentially' is non-truth-conditional (in relevance-theoretic terms, it carries attitudinal information which contributes to higher level explicatures rather than to the proposition expressed). However, sentence adverbials such as 'seriously' and 'confidentially' seem to share conceptual meaning with their manner adverbial counterparts, which contribute to truth-conditional content in the regular way:

(20) I asked him confidentially if he was pleased.

(21) Michael took his role as team captain seriously.

They are thus different from the procedural connectives such as 'so' and 'after all' described above, which have no synonymous truth-conditional counterparts. It therefore seems reasonable to treat 'confidentially' as encoding the same concept in (18) and in (20), but to treat it as contributing to the basic explicature in (20), and to higher level explicatures in (18).

The final logically possible combination involves procedural meaning which constrains the truth-conditional content of an utterance (as opposed to directly encoding a concept which is a constituent of this truth-conditional content). Wilson and Sperber (1993) and Hedley (2007) argue that pronouns can be seen as falling into this category, since they 'guide the search for the intended referent, which is part of the proposition expressed' (Wilson and Sperber 1993, p. 23). Wilson and Sperber (1993), Powell (1998) and Hedley (2007) make reference to Kaplan's 1989 distinction between the content and character of a pronominal. Consider a simple sentence containing a pronoun.

(22) She has a big nose.

The utterance in (22) will be true in just those cases where the person referred to by the pronoun ‘she’ has a big nose. In order to evaluate the truth of this utterance we need to know who ‘she’ refers to. It is not the pronoun itself that appears in the proposition expressed but the referent of the pronoun. As Hedley (2005, p. 44) puts it, ‘their meaning is computational, not representational’. For Kaplan, the ‘content’ of a pronominal is the individual that is contributed to the proposition expressed, and the ‘character’ is the rule by which this content is identified in any given context. In sum, pronouns ‘determine the content (the propositional constituent) for a particular occurrence of an indexical. But they are not a part of the content (they constitute no part of the propositional constituent)’ (Kaplan, 1989, p. 523). When this account is suitably psychologised to fit with a more cognitive approach², it translates into the claim that they are both truth-conditional and procedural.

My analysis of referring expressions is based on the assumption that pronouns are not the only examples of this type. I will argue that other types of referring expressions also encode procedural constraints on explicit content. Although Wilson and Sperber (1993) and Hedley (2007) treat the type of procedural information carried by pronouns as providing constraints on the derivation of the proposition expressed, I believe that the importance of the procedural meaning of referring expressions is underestimated if we focus only on the fact that it guides the hearer to the intended referent. Crucially, the role of referring expressions is to guide the hearer not just to an intended referent, but to an intended overall interpretation, involving each of the three subtasks in (9)-(11). In the next sections, I consider how the choice of referring expression, with its combination of conceptual and procedural information, guides the hearer to the intended explicit content, and how it may play a role in the derivation of contextual assumptions and implications. Whilst Ariel and Gundel et al focus on the contribution of referring expressions to explicit content, I argue that choice of referring expressions cannot be understood fully until we acknowledge and examine their contribution to the other two tasks as well.

² As Blakemore (2002) puts it, ‘an expression which encodes procedural information encodes information which is not a constituent of the conceptual representation over which inferential computations are performed’ (2002, p. 82).

4.2. Explicitly communicated meaning

4.2.1 Determining the proposition expressed

Although I shall be arguing that the role of referring expressions goes beyond the matter of resolving reference, it should not be forgotten that helping the hearer construct a representation of the intended referent that will figure in the proposition expressed is an important part of their function. In order to motivate a relevance theory account that does not rely on Accessibility, Givenness or similar theory-external constructs, I need to show how this takes place. The relevance theoretic account of referring expressions proposed here draws on the conceptual / procedural distinction outlined above and the relevance theoretic comprehension procedure. The way in which the principles of relevance are applied is subtly but crucially different from the way assumed in Ariel's and Gundel's accounts. According to Accessibility theory and the Givenness hierarchy, each referring expression type conventionally (i.e. linguistically) encodes information about the accessibility or cognitive status of the intended referent, which guides the hearer in finding the intended referent amongst the potential referents in the discourse context. For Ariel, referring expressions are 'guidelines for retrievals' (Ariel 1988, p. 68) and for Gundel they 'conventionally signal different cognitive statuses' (Gundel, Hedberg & Zacharski 1993, p. 274). In both cases, it is therefore possible for the speaker's choice of referring expression to direct the hearer to an entity with low accessibility or low cognitive status in preference to higher ones. Ariel clearly states that 'since speakers may wish to refer not to the most salient potential referent, they should overtly instruct the addressee to search for a less salient antecedent' (1994, p. 20). Let us consider this point in relation to the relevance theoretic comprehension procedure, given in (14) above.

The idea that the hearer may ignore or jump over highly accessible potential referents to reach less accessible ones seems to be directly at odds with the comprehension procedure. A hearer following this procedure will always test potential referents and contextual assumptions in order of accessibility, and signalling to the hearer that he should ignore a highly accessible referent in favour of a less accessible one is therefore in direct conflict with a basic principle of relevance theory. If a hearer really were following the relevance theoretic comprehension procedure, then the main effect of such a signal would probably be to cause confusion. However, as Ariel observes, speakers do not always and only refer to the most accessible potential referent in the discourse context. I will argue that we can reconcile this observation with the

relevance theoretic comprehension procedure by re-examining the role of referring expressions. Given that the hearer will automatically follow the comprehension procedure in (14), a speaker who wants to be understood must aim to construct her utterance so that the first interpretation that satisfies the hearer's expectations of relevance is the intended one. In terms of reference resolution, this means that the intended referent should be the most accessible potential referent capable of yielding an overall interpretation that is relevant in the expected way. If the intended referent is not already the most accessible one in the discourse context, the speaker has a large range of potential referring expressions to choose from in order to make this the case. Since, in the framework I propose, referring expressions can encode both conceptual and procedural information, she may exploit both conceptual and procedural information to achieve this end. Logically, there are two strategies available to the speaker. She may use the referring expression to rule out those potential referents which are more accessible than the intended one, or she may use it to increase the accessibility of the intended referent. I will explore both these strategies in the next section, as I consider how both conceptual and procedural meaning may play a part in reference resolution.

In proposing her own, relevance-based account of referring expressions as an alternative to Ariel's and Gundel's approaches, Reboul (1998; 1999) introduces the idea that both conceptual and procedural information may play a role in reference resolution. In her framework, referring expressions are interpreted relative to a 'domain of reference', which she defines as a 'subset of the set of mental representations for a given individual at a given time' (1998, p. 3). A domain of reference is constructed for each new referring expression in the discourse, with its construction constrained by considerations of relevance. According to Reboul, the referring expression itself may provide constraints on the construction of the domain, with both conceptual and procedural information contributing to its construction. Like Wilson and Sperber (1993) and Hedley (2007), she treats pronouns as encoding procedural information, and analyses each pronoun as restricting the domain of reference to just mental representations of a particular gender or number. Similarly, she treats demonstratives as encoding a spatial constraint restricting the domain of reference to either distal or proximal representations. In her framework, conceptual information may also constrain domain construction. For instance, the information encoded by the nominal of a definite description will restrict the domain of reference to just mental representations with appropriate properties. On this account, then, a domain of reference is constructed for each referring expression, and reference will be resolved on one of the mental

representations in that domain. For example, the definite description ‘the black cat’ constrains the domain of reference using both conceptual and procedural information. The definite determiner ‘the’ contributes a procedural constraint which specifies that the domain of reference should be ‘a set Q of objects of which one is N while all the others are not’, and the conceptual information encoded by ‘black cat’ constrains the domain of reference to just those things with the attributes of N: in this case ‘both black and a cat’ (1999, p. 8).

In the account of the use of referring expressions that I present below, I develop Reboul’s ideas, but move away from her notion of a constructed ‘domain of reference’. Instead, I argue that procedural and conceptual information is used to constrain the set of potential referents to a point where the intended referent is the most accessible one in the given discourse context.

4.2.2 Conceptual information and explicit meaning

The conceptual information (if any) encoded by a referring expression will rule out any potential referents that are not compatible with it. Imagine a room full of cats, and consider the many possible ways in which a speaker could refer to a particular cat.

(23) The cat looks hungry.

(24) The black cat looks hungry.

(25) The black cat in the corner looks hungry.

(26) The black cat in the corner with the red collar looks hungry.

In discussing these examples in a cognitive framework such as relevance theory, it is important to distinguish between the actual objects which are the intended referents of the referring expressions, and the mental representations of those objects which will be more or less accessible to speaker and hearer on a given occasion. ‘Accessibility’ in relevance theory is a term that applies to mental representations rather than to actual objects, and I will try to talk systematically of representations as more or less ‘accessible’ and of objects as being more or less ‘salient’ in a given environment.

To refer to a particular cat in the room, the speaker must use a referring expression that makes accessible to the hearer a mental representation of the intended

cat. As argued above, a hearer following the relevance-theoretic comprehension procedure will tentatively select the most accessible representation which is compatible with the linguistically encoded information (and will accept the resulting overall interpretation if it is relevant in the expected way). Thus, utterance (23) would usually be infelicitous in a cat-filled room, since it fails to make a representation of one particular cat more accessible than the others. However, the same utterance would be perfectly acceptable in a room with eight dogs and only one cat. Alternatively, if in the cat-only room, there is one cat that for some reason is more salient than the others - for example, all the cats are asleep, except for one which is worrying for food - then a representation of that cat will be more accessible, reference will be tentatively resolved on it, and (23) will be felicitous if it leads on to an overall interpretation that is relevant in the expected way. If there is no particularly salient cat in the cat-filled room, but only one cat happens to be black, then the speaker may refer to it felicitously using utterance (24). In this case, she uses conceptual information to narrow down the set of potential referents to just black cats; and, since there is only one for which a mental representation is highly accessible, reference will be tentatively resolved on it. Anything not compatible with the conceptual conditions imposed by the referring expression - in this case, being both black and a cat - will be excluded from the set of potential referents, and could therefore be seen as having zero accessibility³. However, utterance (24) is likely to be infelicitous if there are several black cats in the room, since it would fail to make a representation of one black cat more accessible than the others. In this situation, the speaker may be required to further narrow the set, as in (25) or (26), and so on.

Let us compare this account of the role of conceptual information with the analyses offered by Ariel and Gundel. We may refer to a cat using any number of descriptions including 'a cat', 'that cat', and 'the cat'. From a common sense perspective, these referring expressions share the conceptual information encoded by the word 'cat'. Analysing referring expressions as encoding conceptual constraints on the set of potential referents directly captures this common sense intuition in a way that the Accessibility and Givenness scales do not. In both Accessibility theory and the Givenness hierarchy, it is the category of the referring expression (definite description, complex demonstrative, indefinite description, etc) that determines its place on the scale, and hence the contribution it makes to reference resolution. Neither account has

³ It is worth noting here that the encoded conceptual material is merely a clue to the intended referent. A metaphorical referring expression may narrow down the set of potential referents without necessarily excluding anything that doesn't satisfy the literally encoded concept.

anything to say about the role that conceptual content may play in reference resolution. Furthermore, Accessibility theory distinguishes only between ‘long’ and ‘short’ definite descriptions, whilst the Givenness hierarchy treats them all equally as indicating the cognitive status of being ‘uniquely identifiable’. My approach has the advantage of offering a potential explanation for how the speaker chooses among the indefinite range of options available when constructing a definite description, and it makes clear predictions about which will be felicitous in certain discourse contexts, and which will not. In effect, the discourse context can be viewed as including a set of potential referents, each of which can be mentally represented in a variety of more or less accessible ways, and the referring expression can be seen as a means by which the speaker may select a subset of the potential referents, such that a representation of the intended referent is the most accessible to the hearer in that subset.

4.2.3 Procedural information and explicit meaning

Although the conceptual information carried by descriptive phrases plays a crucial role in reference resolution, it is not the only means by which a speaker can constrain the set of potential referents. In a framework with both conceptual and procedural meaning, referring expressions may also encode procedural constraints. In section 4.1.3, I mentioned some existing analyses of pronouns as encoding procedural information that constrains the truth conditional content of an utterance (e.g. ‘you’ constrains the class of potential referents to sets of individuals that include the addressee). Here I want to suggest that the demonstratives and determiners that figure in more complex referring expressions play a similar role.

Perhaps the easiest way to illustrate this point is by considering the role of distal and proximal determiners. Imagine a scenario in which Emily is sitting in a tea shop when the waitress brings a trolley with a large cake on it. The waitress offers Emily a slice of the cake, and also explains that if she prefers, she may have a slice of the cake on display in the shop window. As it happens, the cake on the trolley is more to Emily’s taste. Now consider how the following utterances by Emily might be understood:

(27) I’ll have a slice of the cake.

(28) I’ll have a slice of this cake.

(29) I'll have a slice of the cake on the trolley in front of me.

Although utterances (27) and (28) encode the same conceptual information, it is likely that reference resolution will succeed in (28) but fail in (27). In both cases, the conceptually encoded meaning ('slice of cake') narrows the set of potential referents to two— a slice of the cake on the trolley, and a slice of the cake in the window. Utterance (28) succeeds because the procedural information encoded in the determiner narrows the set of potential referents still further, to a point where it includes only a slice of the cake on the trolley.⁴ Utterance (29) would probably lead to reference being resolved on the same slice of cake, but would normally be ruled out on effort grounds. The processing of the extra conceptual information in (29) would put the hearer to extra processing effort without achieving any extra effects (unless the waitress has given evidence of being particularly hostile or slow on the uptake).

I am therefore suggesting that referring expressions (and in particular pronouns and determiners) may encode procedural information which can narrow the set of potential referents more efficiently than would be achieved by encoding further conceptual constraints. In following the sections I will consider in more detail what this procedural information might look like in specific cases.

In certain cases, it seems that the conceptual information encoded by a referring expression is enough on its own to narrow the field of candidate referents to a point where reference resolution can be successfully achieved, and yet we still find a contrast in acceptability or interpretation between referring forms which differ only in the determiner used. I will return to this point in section 4.3.2, where I will argue that in these cases, the procedural information carried by the determiner contributes not to reference resolution but to the implicit phase of comprehension, and guides the hearer as to the type of inferences he should draw.

According to Ariel (1994), a major problem with pragmatic accounts of reference resolution is their inability to 'explain to us how one decides among various potential antecedents' (1994, p. 11). However, the relevance theoretic comprehension procedure does just this. Any hearer following the procedure will test representations of potential referents in order of accessibility, and stop when his expectation of relevance is satisfied. The conceptual and procedural information encoded by the referring

⁴ In section 5.2 I give a more detailed account of what exactly this procedural information might be. However, for present purposes it is enough to follow the suggestion of Reboul (1997) that the information these demonstratives encode 'has to do with the relative distance between the speaker and object designated', and that it 'has a spatial content which closely parallels that between "here" and "there"' (1997, p. 11)

expression will constrain the set of potential referents, and it is this constraining role that drives the speaker's choice of a particular referring expression. Thus, rather than uniquely indicating to the hearer which potential referent should be chosen, referring expressions narrow the set of potential referents to a point where the relevance-theoretic comprehension procedure can successfully complete the task. As with all cases where semantics and pragmatics interact, the kind and extent of constraint required will vary with the context, and so the same entity may be referred to in several different ways in different circumstances.

4.2.4 Applying the constraints

Assuming a model in which conceptual and procedural constraints interact to guide the hearer towards the intended interpretation, I will now re-examine the problematic examples from (1)-(4) above. In this framework, the acceptability of a given referring expression should be predictable on the basis of the encoded conceptual and procedural information and the accessibility of a representation of the intended referent relative to those of other potential referents.

The version which Gundel and Mulhern interpret as conveying that 'these primitive reptiles' is coreferential with the set of all parietosaurs is repeated here for convenience:

- (1) A restudy of parietosaurs reveals that these primitive reptiles are the nearest relatives of turtles.

To give a full account of the interpretation of this example, it will be necessary to consider the complicated interaction between reference resolution, context construction and implicated conclusions. In this section, I begin by looking at how encoded conceptual and procedural information contributes to the retrieval of the proposition expressed. I will argue that the proposition expressed by (1) is the same as that in (2) and the proposition expressed by (3) is the same as that in (4). My suggestion is that the apparent difference in the acceptability of the utterances is a result of interaction between the proposition expressed and accessible contextual assumptions, on the one hand, and of the way the procedural information guides the inferential process, on the other. First, however, I will consider how the explicit content is derived.

The referring expressions in (1) and (2) both encode the conceptual information 'primitive reptiles', which narrows the set of potential referents to include only (sets of)

things that are both primitive and reptiles. Given that there is one highly salient group of primitive reptiles in the discourse context, it might be tempting to suggest that this is all the information the speaker needs to provide in order to convey her intended meaning. However, consider the utterance in (30).

- (30) A restudy of pareiasaurs reveals that primitive reptiles are the nearest relatives of turtles.

When a bare-NP is used instead of a definite-NP, the interpretation changes significantly. The NP ‘primitive reptiles’ receives a generic interpretation on which it picks out primitive reptiles as a kind (as opposed to pareiasaurs as a subset of primitive reptiles). (30) will be true if and only if members of the kind ‘primitive reptiles’ are the nearest relatives of turtles. Contrast this with the interpretation of (1) (and arguably (2), on the anaphoric epithet reading discussed above), where the NP ‘these primitive reptiles’ is co-referential with the NP ‘pareiasaurs’. On this interpretation (1) is true only if pareiasaurs are the nearest relatives of turtles, and asserts nothing about primitive reptiles in general. Thus, the inclusion of the demonstrative determiner in (1) affects the truth conditional content of the utterance.

As with most procedural information, the procedural information carried by the determiner can be difficult to paraphrase in conceptual terms. According to Wilson and Sperber (1993, p. 17) ‘Conceptual representations can be brought to consciousness: procedures cannot’. More generally, conceptual information encoded by familiar lexical items is relatively easy to grasp and paraphrase. However, linguistic elements that encode procedural meaning can be notoriously difficult to paraphrase, and the role they play extremely difficult to pin down. Still, in the case of (1), the procedural information carried by the demonstrative determiner in the phrase ‘these primitive reptiles’ seems to divert the hearer from a generic interpretation, towards one in which a definite group of primitive reptiles is picked out. In a later section, I will consider why procedural information in general is so hard to paraphrase.

As Gundel (2003, p. 130) notes, the definite description ‘the primitive reptiles’ in (2) could logically pick out ‘the whole class of primitive reptiles’. Why, then, is (2) judged to be unacceptable or difficult to process on this reading, rather than simply being interpreted as a paraphrase of (30)? I suggest that the answer to this question follows from the relevance theory claim that a speaker aiming at optimal relevance will not deliberately put the hearer to any gratuitous processing effort. The conceptual

information carried by (30) narrows the set of potential referents to just those which are both primitive and reptiles, and in the absence of a definite determiner ('the' or 'these'), the hearer settles on the generic reading.⁵ Although the same interpretation could in principle be achieved by adding the definite determiner and intending the hearer to resolve reference on the definite group of all primitive reptiles, a speaker who chose this formulation would be demanding extra processing effort from the hearer, but failing to offer any extra reward. Moreover, if the speaker wishes to refer to all primitive reptiles, then she runs less risk of misunderstanding if she uses the bare-NP form. Whilst the bare-NP unequivocally picks out the set of all primitive reptiles, there are many different subsets that the definite description could potentially pick out. If (2) is deemed acceptable at all, then it is under the anaphoric epithet interpretation where the explicit content is parallel with (1), and not with (30).

This brings us back to the issue of how the hearer of (1) identifies the particular group of primitive reptiles intended by the speaker. Whilst pareiasaurs are a highly salient group of primitive reptiles, it could be possible to identify various other subsets within the set of primitive reptiles, for example crocodylians, which might also fit the conceptual constraints. If there are several logically possible groups of primitive reptiles that could satisfy the referential constraints, then how does the hearer settle on the interpretation where the definite NP is co-referential with the pareiasaurs?

I suggest that the answer to this question follows from the relevance theoretic comprehension procedure. Recall that a hearer following this procedure will test potential referents, and the resulting overall interpretations, in order of accessibility. We therefore predict that the hearer will first test the most accessible referent that is not excluded from the set of potential referents by the information encoded by this referring expression. The speaker has used conceptual information to narrow the set of potential referents to groups of primitive reptiles, and the procedural information carried by the determiner narrows the set further to include only definite (i.e. identifiable) groups of primitive reptiles. Whilst other possibilities are logically available to the hearer, the subset of pareiasaurs is clearly the most accessible one in the context for both (1) and (2). The hearer will therefore test the hypothesis that 'pareiasaurs' and 'the/these primitive reptiles' are co-referential, and, on finding that this interpretation yields enough implications to satisfy his expectation of relevance, will look no further. A similar process is followed in the interpretation of (3) and (4) ('the two groups'/'these

⁵ It has been proposed that there is some sort of phonologically null generic operator in such cases. (see Papafragou (1996) for discussion)

two groups'), and the hearer resolves reference on the two-group set of pareasosaurs and turtles.

Having established that some sort of definite determiner is required in these cases, we reach the crux of Gundel and Mulkern's argument against a purely pragmatic account, and their justification for introducing their scale of Givenness. Although my account so far explains the contribution of both conceptual and procedural information to the proposition expressed, the contrast between 'the primitive reptiles' and 'these primitive reptiles' remains unexplained. According to Gundel and Mulkern, such an explanation is beyond the scope of relevance theory alone, and indeed, I have so far provided no reason why a speaker might prefer one version over the other. In the following sections, I will address this issue as I move on from the explicit phase of comprehension to discuss the role that both conceptual and procedural information may play during the implicit phase. Acknowledging this role, which is largely overlooked in the Accessibility and Givenness accounts, will allow us to develop an analysis which captures the subtle differences in acceptability between the utterances in (1)-(4). I will argue that Gundel and Mulkern are wrong to characterise the contrast between (2) and (3) as lying purely on the explicit side of comprehension. In their view, the various forms of referring expression contribute to the process of reference resolution, and therefore to the proposition expressed. I hope to have shown in this section that the propositions expressed by (1) and (2), and by (3) and (4), are the same. In all cases, the hearer narrows down the set of potential referents to just those which are sets of things that satisfy the conceptual and procedural constraints. The differences in acceptability result from interaction between the propositions expressed by the utterances and an accessible context, guided by the relevance-theoretic comprehension procedure. I propose that the same processes and constraints are in play in (3) ('the two groups') as in (2) ('the primitive reptiles'), but that considerations of relevance result in the infelicity of (2), and not of (3).

My claim is that the procedural information carried by determiners can not only contribute to the proposition expressed, but can also affect the inferences drawn during the implicit phase of comprehension. I will return to these examples in the next section, when I consider how referring expressions contribute to the implicit side of utterance interpretation. However, first I will discuss how the conceptual content of a referring expression may contribute to implicit meaning.

4.3 Beyond reference

4.3.1 Conceptual information and implicit meaning

We have seen how conceptual information can affect the derivation of the explicit content of an utterance by guiding the hearer to the intended referent. However, derivation of the explicit content is just one of the sub-tasks involved in utterance interpretation. In order to understand what is implicitly communicated, the hearer has to construct ‘an appropriate hypothesis about the intended contextual assumptions’ and ‘an appropriate hypothesis about the intended contextual implications’ (Wilson & Sperber, 2004, p. 261). For the purposes of the discussion to follow, it is important to stress that according to relevance theory, all three sub-tasks take place in parallel, and the resulting hypotheses may be ‘revised or elaborated as the utterance unfolds’ (ibid). Although I discuss each process separately for clarity and simplicity, they should be seen as interrelated sub-tasks ‘embedded within the overall process of constructing a hypothesis about the speaker’s meaning’ (ibid, p.262).

To illustrate the contribution that conceptual information can make to implicatures, I return once again to the examples in (1) and (2) discussed by Gundel and Mulhern (1998) and Gundel (2003). Compare (1) (‘A study of pareiasaurs reveals that these primitive reptiles are the nearest relatives of turtles’) with the utterance in (31).

(31) A restudy of pareiasaurs reveals that *they* are the nearest relatives of turtles.

In both (1) and (31), it is likely that the hearer will resolve the reference of ‘these primitive reptiles’ and ‘they’ in the same way: in both cases, the referring expression will be interpreted as co-referential with the antecedent phrase ‘pareiasaurs’. It therefore seems that, despite the discussion in 4.2.2 above, the conceptual information ‘primitive reptiles’ is unnecessary as far as resolving reference is concerned, since the pronoun ‘they’, which does not carry this information, works just as well. Indeed, I predict that a hearer could correctly resolve reference in (31) even if he did not know what a pareiasaur was, and had only limited knowledge of primitive reptiles. This falls out naturally from my earlier discussion, since we have seen that the set of pareiasaurs is the most accessible potential referent, and so will be tested first.

However, if the conceptual constraints discussed above are not strictly necessary for reference resolution, then why is (1) - which is linguistically more complex than (31) - not ruled out by considerations of effort? Once again, the answer is provided by

the relevance theoretic comprehension procedure and the associated principles of relevance. According to the procedure, the hearer should test interpretive hypotheses in order of accessibility, and stop when his expectation of relevance is satisfied. So far I have concentrated on the first clause of the procedure, which stipulates that interpretations will be tested in order of accessibility, but the second clause is equally important and it will help us gain insight in this case. The hearer can presume that the utterance will be optimally relevant, where optimal relevance is defined as in (13) above. Once again, we must return to the basics of relevance theory to remind ourselves what makes an input relevant enough to be worth processing. The interpretation of (1), as compared with (31), involves the same assignment of reference to the referring expression, but demands that the hearer process more linguistic material. He must also narrow the set of potential referents to just those things that are both primitive and reptiles, even though the pareiasaurs are already the most salient potential referent in the discourse context. This extra effort will only be justified if it is rewarded with extra cognitive effects. I claim that that is exactly what happens in this case, and that the extra effects that the speaker intends to convey explain the alternative choices of referring expression.

Recall the three types of cognitive effect that contribute to relevance, as discussed in chapter 2 above. The processing of an input in a given context may lead to the strengthening of a contextual assumption, it may contradict and eliminate a contextual assumption, or it may combine with contextual assumptions to yield implications that follow from the input and the context together, but from neither the input nor the context on its own. The greater the effects, and the smaller the effort required to derive these effects, the more relevant the input will be (for that person, at that time).

The effects that will be derived from the additional conceptual information in the referring expression ‘these primitive reptiles’ in (1) depend on the contextual assumptions available to the hearer. Consider the following possible contextual assumptions and their likely interaction with the input in (1).

(32) Pareiasaurs are primitive reptiles.

(33) Pareiasaurs are primitive mammals.

If the hearer of (1) holds the assumption in (32), then the utterance will have the effect of strengthening this assumption. The size of the effect will depend on the strength with

which the hearer held (32) to begin with, and his judgement of the speaker's expertise and honesty. If the hearer held assumption (33), by contrast, then processing (1) will lead to the contradiction and possible elimination of this assumption. Once again, the size of the effect will depend on the hearer's assessment of the speaker's abilities and honesty (not everything that is communicated is believed). Both utterances make it possible to derive cognitive effects that would not have been derived from the utterance in (1).

Now consider a third possibility. The hearer holds no assumptions about pareiasaurs: he has never heard of them and has no idea what they are. He will still process the utterance in (1) with the expectation that it will be optimally relevant. Following the relevance theoretic comprehension procedure, he will test out representations of possible referents in order of accessibility. In this case, since he has no full-fledged concept of pariesaur, he starts, (as children do when processing unfamiliar words) from a metarepresentation of the full-fledged concept that he takes the speaker to possess. Recall, as part of the interpretation process, the hearer must construct 'an appropriate hypothesis about the intended contextual assumptions (in relevance-theoretic terms, implicated premises)' (Wilson & Sperber, 2004, p. 261). An easily accessible contextual assumption in this case would be (34).

(34) 'Pareiasaurs' are primitive reptiles.

This implicated premise is likely to be relevant to the hearer in its own right, as a new piece of information that may be useful in the processing of later utterances.⁶ So (1) leads to a cognitive effect that would not have been derived if the utterance in (31) had been used in its place. The extra effort of processing a complex demonstrative rather than a pronoun is offset by the extra rewards of accessing the implicature in (34).

The relevance theoretic approach to communication stresses the importance of interpreting utterances in a context. When Gundel (2003) and Gundel and Mulhern (1998) discuss the example in (1), they do so without discussing the discourse context in which it occurs. In fact, example (1) comes from a scientific paper (Lee, 1993), and I suggest that the author has formulated his utterance so as to produce as many cognitive effects in as many readers as possible, as efficiently as possible. Different readers will

⁶ In many ways, (34) functions similarly to a bridging assumption. However, I follow the view of Matsui (2000) that this is not a genuine bridging assumption since there is an "explicitly mentioned antecedent in the previous discourse" (2000, p.20).

hold different assumptions, perhaps including those in (32) or (33), and so the potential for deriving the cognitive effects discussed above is clear.

However, I suggest that this kind of carefully formulated textbook or academic style is not the only way in which conceptual information can affect the implicit side of communication. Consider the utterances in (35) and (36).

(35) John_i went into town for his lunch. He_i was late back to the office.

(36) John_i went into town for his lunch. The rascal_i was late back to the office.

In (35), the effort required from the hearer in narrowing of the set of potential referents is minimal. He need only exclude potential referents who are not male⁷. When processing (36), however, the hearer is required to do more work. The referring expression 'the rascal' limits the set of potential referents to those the speaker considers to have the property of being a rascal. Logically, the hearer has two choices at this stage: he can either treat John as a member of this set, or assume he is not a member. Taking the second option would leave him with nothing on which to resolve reference, and his expectation that the utterance will be optimally relevant would be disappointed. He is therefore likely to try the other option and construct the contextual assumption in (37).

(37) The speaker thinks that John is a rascal.

Reference is again resolved on John, since he remains the only potential referent whose representation is accessible enough. However, the extra effort required by this formulation is rewarded with extra effects: the hearer now knows that the speaker regards John as a rascal. This may be relevant in its own right as a piece of new information, or it might combine with other existing contextual assumptions to increase the contextual implications of (36): for instance, it might combine with the assumption in (38), to yield the contextual implication in (39).

(38) To call someone a rascal is to condemn some aspect of their behaviour.

⁷ It is not clear that even this level of restriction is necessary here, since there are no non-male potential referents to be ruled out. In chapter 6 I discuss similar examples, and consider the significance of gender marking on pronouns in more detail.

(39) The speaker is condemning John's returning late from lunch.

I suggest that, in practice, the use of a referring expression such as 'the rascal' is likely to lead to a whole range of weak implicatures based on assumptions about the nature of rascals, and the speaker's opinion of John. The speaker has therefore produced a wide range of cognitive effects for only a little extra effort. In this way, the conceptually encoded content of the referring expression has contributed to the inferential phase of communication, and we can see referring expressions as just one more device by which a speaker can achieve optimal relevance.

However, there is no guarantee that adding extra conceptual content will always result in more or different cognitive effects. Consider (40).

(40) ??John_i went into town for his lunch. The man_i was late back to the office.

The structures in (36) and (40) take the same form. Both use definite descriptions ('the rascal'/'the man') to refer back to John. However, the result is less acceptable in (40) than in (36)⁸. The difference in acceptability of (36) and (40) lies in the relation between the processing effort the referring expressions require and the effects that they offer in return. The narrowing required by the pronoun 'he' in (35) is minimal. Although the referring expression 'the rascal' in (36) requires more narrowing, this is rewarded with extra cognitive effects. In (40), the expression 'the man' requires more effort, but does not provide any obvious extra effects. The hearer is required to narrow the set of potential referents to a sub-set which qualify as 'men', but given that John is the only potential referent in the discourse context, this is an apparently gratuitous extra layer of narrowing, which calls for extra processing effort, and should therefore offer extra effects. However, there are no obvious cognitive effects that can result from choosing the referring expression 'the man'. The only conceivable contextual assumption with which the input could combine is that in (41).

(41) John is a man.

But unless the hearer was under the misconception that John might be a dog or other type of non-human, or perhaps a female, it is unlikely that the implicated premise in

⁸ Neither Accessibility theory nor the GHZ framework has anything to say about why this should be. Both treat definite descriptions as a unitary set, whose members should be acceptable under the same conditions.

(41) would contribute to relevance, either by strengthening existing assumptions, by contradicting or eliminating existing assumptions, or by combining with existing assumptions to yield contextual implications.⁹ Thus, the extra effort required by (40) is gratuitous, and the result should be stylistically less than optimal.

4.3.2 Procedural information and implicit meaning

So far, I have discussed how the conceptual information encoded by referring expressions contributes to explicit and implicit content, and how the procedural information encoded by pronouns and determiners can affect the proposition expressed, and therefore contribute to explicit content. In this section, I examine the fourth logically possible combination: the effect of this procedural information on the implicit side of utterance interpretation. I return to the examples in (1)-(4), and the differences in acceptability of the alternations in determiner forms.

In section 4.2.3, I showed how the spatial information encoded by the determiner in a complex demonstrative may exclude certain potential referents and thus help the hearer to resolve reference. As discussed above, this extra layer of narrowing does not seem to be necessary for reference resolution in the case of (1). In both (1) ('A restudy of pareiasaurs reveals that these primitive reptiles are the nearest relative of turtles') and (2) ('A restudy of pareiasaurs reveals that the primitive reptiles are the nearest relative of turtles') 'pareiasaurs' is the most accessible potential referent for the referring expression 'these/the primitive reptiles', and in both cases reference will be resolved on it. However, this is not to say that the indication of proximity carried by the determiner in (1) plays no role in the interpretation of this utterance. In processing (1), the hearer will still presume that it is optimally relevant, and is therefore entitled to expect that the speaker will not deliberately cause him gratuitous effort. Since the procedural information encoded by the complex demonstrative is not necessary for deriving the explicit content, he should expect some other effect to justify its use.

In general, indications of proximity are only appropriate when the spatial location of the referent is significant in some way, for example, by distinguishing it from some other non-proximal entity. Thus, use of the complex demonstrative form 'this/these N' indicates that the intended referent contrasts with some other entity of the same type. Both Reboul (1997) and Powell (2002) offer some suggestions about how to capture the difference between complex demonstratives and definite descriptions, and both seem to

⁹ It is likely that the hearer already holds the assumption that John is a man with sufficient strength that any further strengthening will not be possible or significant.

be appealing to this idea of contrast. I believe it does indeed shed some light on the procedures encoded by these forms of referring expression.

As discussed above, Reboul considers the procedural information encoded by the definite article ‘the’, and suggests that it restricts interpretation to ‘a set Q of objects of which one is N while all the others are not’ (1999, p. 8). Here, N corresponds to the conceptual information encoded by the definite description accompanying the determiner. Powell (2002) discusses the difference between definite descriptions and complex demonstratives as part of his work on reference. Complex demonstratives, he claims, are ‘communicative tools designed for a particular purpose, that purpose being to talk about particular members of the nominal class’¹⁰ (2002, p. 226). He goes on to discuss how complex demonstratives contrast with definite descriptions: ‘the nominal of a definite description must be uniquely denoting, that of a complex demonstrative must not be uniquely denoting.’ (2002, p. 230), and describes their procedural meaning as follows:

whereas a definite description ‘the F’ exploits, by dint of its linguistic meaning, the property of being a unique F in order to guide a hearer to the intended interpretation, a complex demonstrative ‘that F’ exploits the property of being a non-unique F to guide the hearer to an individual concept (2002, p. 230).

Powell places particular emphasis on the idea that the object picked out by the descriptive content of a complex demonstrative is non-unique, and the referent is therefore picked out as one instance of an F in contrast to all the other instances of Fs.¹¹

I argued in 4.3.1 that for a hearer following the relevance theoretic comprehension procedure, the procedural information about spatial relations encoded by a complex demonstrative may be superfluous during the explicit phase of communication. I now want to suggest that this information can be exploited during the implicit phase to constrain the type of inferences that the hearer is encouraged to draw. The contrast in appropriateness between the resulting inferences should help to explain the differences in acceptability between the utterances in (1) - (4). We must therefore ask what role this procedural information plays in the interpretation process, and exactly how the

¹⁰ By nominal class, Powell seems to mean the set of things that fall under the concept encoded by the complex demonstrative.

¹¹ Again, I will discuss this notion of ‘contrast’ in relation to demonstratives in more detail in section 5.2 below.

differences in acceptability of utterances (1) - (4), and the related examples in (5)-(8), might be explained.

As discussed above, a hearer following the comprehension procedure tests interpretive hypotheses in order of accessibility. Given the context made accessible by processing the previous utterance (1) ('A restudy of pareiasaurs reveals that these primitive reptiles are the nearest relatives of turtles'), the most salient set of two groups for the interpretation of (3) and (4) should consist of the pareiasaurs and the turtles. The conceptual information provided by the phrase 'the/these two groups' should be enough to ensure that the hearer retrieves the intended referent, just as it was in (1).

Recall my claim that the procedural information about spatial relations encoded by a complex demonstrative is only relevant if there is another potential referent in a different spatial relation to the speaker. By encoding such information, as in (4) ('These two groups...'), the speaker therefore creates a contrast effect. No such effect is created by use of the uniquely-denoting definite description form 'the two groups', as in (3). Thus, in (4), the two groups under discussion are presented as contrasting with other potential groups. This in turn gives rise to a range of weak implicatures about other groups that do not share the same characteristics. The same implicatures are not derived in (3), where the only two groups of any relevance are the turtles and the pareiasaurs.

On this account, the contrast in acceptability between (2) and (3) results not from any encoded level of Accessibility or Givenness, but from the appropriateness of the inferences derivable from each version. Both (4), with its indication of contrast, and (3), without it, are equally acceptable in the discourse context. However, things are different with (1) and (2). As argued above, in (1) 'these primitive reptiles' guides the hearer to an interpretation where the pareiasaurs are set in contrast to other potential groups of primitive reptiles. This is just what the speaker of (1) intends: it is these reptiles, and not others, that are claimed to be the closest relatives of turtles. Logically, only one set of primitive reptiles can be 'the closest relatives of turtles', and thus, an utterance where the intended referent is set in contrast to others is the most appropriate. Notice that if we take away the superlative, then the 'the/these' variants become equally acceptable, as in (5) -(8). Add the superlative back in, and we find the same contrast arising again, with only a marginal anaphoric epithet reading available, as in (42)-(43):

- (42) A restudy of pareiasaurs reveals that *these primitive reptiles* were the largest in Africa.

(43) ??A restudy of pareiasaurs reveals that *the primitive reptiles* were the largest in Africa.

Thus, the role of the demonstrative determiner and the procedural information that it contributes to the interpretation of the complex demonstrative is two-fold. It may help to narrow the set of potential referents and guide the hearer to the explicit meaning by encoding a procedure which further narrows the set of potential referents, or it may guide the hearer in the inferential phase of comprehension. When the resulting inferences conflict with the sense of the utterance, an infelicity effect may occur.

Once again, it is vital to remember that as determination of explicit and implicit content takes place in parallel, the procedural information conveyed by pronouns and determiners is ultimately a device by which a speaker can achieve optimal relevance. Furthermore, the marginality of the judgement on (2)('A restudy of pareiasaurs reveals that the primitive reptiles are the nearest relatives of turtles') seems to suggest that an inferential account rather than a pure decoding account may be appropriate. The fact that the acceptability of these utterances can be affected by the context in which they are uttered is to be expected if the crucial factor is the role they play in what is implicitly communicated.

4.4 Concluding remarks

In this chapter, I have outlined a relevance-based account of referring expressions in which the form of the expression is a guide to the hearer in reference resolution and different forms may yield different inferential effects. I have argued that referring expressions encode conceptual and/or procedural information which may contribute to what is explicitly and/or implicitly communicated. I have argued that this approach provides an explanation for the acceptability judgements on (1)-(4), and this, in turn, weakens the case for introducing theory-external notions of Accessibility or Givenness. In the next chapter, I apply the same approach to the pairs of referring expressions which, according to Accessibility theory, could not be distinguished on a relevance-only account.

Chapter 5: Motivating the distinctions: demonstratives and stressed pronouns

5.1 Introduction

In chapter 4, I outlined a relevance theory approach to referring expressions which uses the conceptual /procedural distinction to show how referring expressions might contribute to what is explicitly and implicitly communicated by an utterance. I worked through the examples given by Gundel and Mulhern (1998) to illustrate differences in interpretation and acceptability in the use of definite descriptions and complex demonstratives, and in doing so, I appealed to the idea that complex demonstratives encode information about proximity and convey suggestions of contrast.

In this chapter, I work out these ideas in more detail and use them to analyse a range of examples which Ariel sees as problematic for a relevance-only account. In section 5.2, I consider the distinction between the proximal and distal determiners ‘this’ and ‘that’. I give a brief outline of how the contribution of demonstratives to reference resolution has been characterised in previous analyses, emphasising the variety of roles they play. I then consider in more detail what distinguishes ‘this’ and ‘that’ from each other, and from the definite determiner ‘the’. In section 5.3, I take a similar approach to the differences between stressed and unstressed pronouns.

In line with the relevance-theoretic approach outlined in chapter 4, I treat referring expressions as encoding conceptual and/or procedural meaning which contributes to the speaker’s overall aim of making her utterance optimally relevant, and assume that in doing so, they may contribute either to reference resolution itself or to the derivation of implicit inferential effects¹. One advantage of this approach is that it allows us to look

¹ There has been some debate about whether complex demonstratives are referential or quantificational. Powell (2001; 2002) gives a detailed overview of the various positions, and argues that ‘complex demonstratives can give rise to genuinely referential or genuinely quantificational truth conditions, according to speaker intention’ (2001, p. 69). In this chapter, I follow Powell in assuming that ‘complex demonstratives are tools used by speakers to indicate that the thought they intend to express contains an individual concept’ (2001, p.67). On this account, those cases which are typically cited as examples of quantificational use, for example (i) and (ii), are analysed as communicating descriptive individual concepts.

- (i) That hominid who discovered how to start fires was a genius.
- (ii) Every father dreads that moment when his eldest child leaves home.

As far as the genuinely referential cases go, ‘referentially-used complex demonstratives contribute nothing but their referents to propositional content’ (2001, p. 71).

beyond the process of reference resolution to consider the contribution of referring expressions to other types of effect: for instance, those associated with ‘emotional’ or ‘metaphorical’ uses of demonstratives mentioned briefly in section 3.3.5.1, which have largely been ignored in the formal literature. I briefly reconsider these so-called ‘special uses’ of referring expressions, and conclude that they are not so special after all.

5.2 Demonstrative determiners

5.2.1 Existing analyses

Various attempts have been made to provide a taxonomy of demonstrative terms and their uses. Following Diessel’s (1999) in-depth cross-linguistic study, Levinson (2004) proposes the classification in figure 1, and illustrates each category with the examples in (1)-(8).

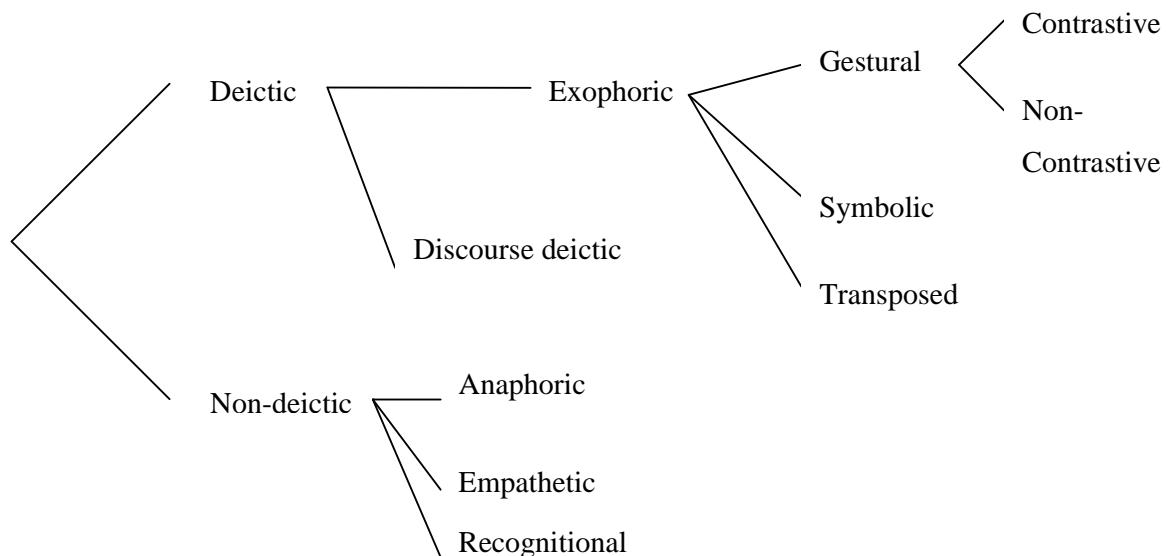


Figure 1: A taxonomy of demonstrative terms (Levinson 2004, p. 108).

- (1) I hurt this finger. (gestural exophoric: requires gesture or presentation of finger)
- (2) I broke this tooth first and then that one next. (gestural contrastive)

In sum, this account of the semantics of complex demonstratives ‘takes a single semantics, a lexically encoded meaning which constrains the mapping from linguistics to conceptual representation for all complex demonstratives, and leaves the rest, i.e....whether the intended individual concept is *de re* or descriptive etc., to pragmatics’ (2001, p.70).

- (3) I like this city. (exophoric symbolic: does not require gesture).
- (4) He looked down and saw the gun: this was the murder weapon, he realized.
(transposed)
- (5) “You are wrong” That’s exactly what she said. (discourse deictic)
- (6) The cowboy entered. This man was not someone to mess with. (anaphoric)
- (7) He went and hit that bastard. (empathetic)
- (8) Do you remember that holiday we spent in the rain in Devon. (recognitional)

Levinson acknowledges that ‘the relations between these uses are probably more complex than this taxonomy suggests’, and that there are many ‘borderline examples’ and ‘fuzzy borders’ (Levinson, 2004, p. 107). Both Diessel and Levinson draw on earlier work by Lakoff (1974) and Fillmore (1982; 1997), who propose similar distinctions, discuss the various categories in more detail and point out many subtleties in the interpretation of demonstratives in English which are of particular interest for pragmatics. I will therefore base my discussion largely on Lakoff and Fillmore’s accounts, bringing in the developments proposed by Diessel and Levinson as appropriate.

Both Lakoff and Fillmore distinguish three broad categories of use which cross-cut Levinson’s distinctions in figure 1: Fillmore identifies ‘gestural’, ‘anaphoric’ and ‘symbolic’ uses, while Lakoff identifies ‘spatio-temporal’, ‘discourse’ and ‘emotional’ uses. Since the two sets of distinctions are roughly parallel, I will consider them together, and use them as the basis for my discussion.

5.2.1.1 Gestural / spatio-temporal uses

This is perhaps the most intuitive of the categories. Fillmore describes gestural uses as those which ‘can be properly interpreted only by somebody who is monitoring some physical aspect of the communication situation’ (1997, p. 62). Some physical demonstration of the intended referent, by gesture, eye-gaze, head nod or so on, must accompany the utterance for the use to be felicitous. As Fillmore puts it, ‘you will expect the word to be accompanied by a gesture or demonstration of some sort,’ (1997,

p. 63). Lakoff (1974, p. 345) defines her first category as comprising ‘literal “pointing” words’, where ‘this’ is used when the object is close to the speaker and ‘that’ when it is further away, ‘particularly when contrasted with another closer’ entity. Although Lakoff is not as explicit as Fillmore about the need for a demonstration to accompany the utterance, she does specify that this type of use ‘normally refers only to items literally in front of the speaker as he speaks’ (Lakoff, 1974, p. 346). Consider (9), uttered by a customer in a bakery. For the utterance to be felicitous, the objects referred to must be physically present and in some way indicated by the speaker (e.g. by pointing, eye gaze, head nod, etc.).

(9) I’ll take this loaf, that sausage roll and three of those buns.

Such cases would, be considered ‘gestural’ by Fillmore and ‘spatio-temporal’ by Lakoff. Whilst ‘spatial’ uses seem straightforward enough under Lakoff’s taxonomy, it remains unclear what she intends by the associated notion of a ‘temporal’ use. She gives no explicit examples of temporal uses in discussing the ‘spatio-temporal’ category, and it is unclear how such uses would differ from the discourse-deictic uses described in the next sub-section

Diessel (1999) classifies these ‘gestural’ and ‘spatial’ uses as ‘exophoric’, and argues that they have three distinctive features which set them apart from other uses:

first, they involve the speaker (or some other person) as the deictic center, second they indicate a deictic contrast on a distance scale...and third, they are often accompanied by a pointing gesture (Diessel, 1999, p. 94).

5.2.1.2 Anaphoric / discourse uses

Demonstrative determiners may be used to refer back or forward in the discourse, and both Fillmore and Lakoff consider examples of this type. Fillmore (1997) does not limit his category of ‘anaphoric’ uses to just those whose referents are explicitly mentioned in the discourse, but includes any ‘happening which is observable at the same time by encoder and decoder’ (1997, p.104). So, according to Fillmore, a speaker might use ‘this’ ‘to refer forward in time to an event or happening, as in (10), or she might use it to introduce a subsequent part of her contribution to the discourse, as in (11).

(10) This is my imitation of a frog.

(11) This is my explanation.

Fillmore draws a distinction here between backward looking reference (anaphora) and forward looking reference (cataphora), and makes the general claim that the distal determiner 'that' is usually associated with anaphora, as in (12), whilst 'this' is used for cataphora, as in (13).

(12) Eat your greens: that's the solution to your problems.

(13) This is the solution to your problems: eat your greens.

The distinction in acceptability between forward and backward-looking anaphora is complicated by interaction with tense and aspect. As Fillmore (1997, p.105) points out, (14) and (15) are, on the whole, more acceptable than (16) and (17), although both involve backward reference. It seems that the perfect aspect and the perfect continuous aspect affect the choice between proximal and distal demonstrative independently of whether the reference is backward or forward looking. I will return to this observation later.

(14) This has been an interesting course.

(15) That was a brilliant lecture.

(16) ?This was an interesting course.²

(17) ?That has been a brilliant lecture.

Lakoff makes a similar observation, pointing out that whilst 'this' can refer both forwards and backwards, 'that' is only available for backwards reference. Again, the distinction between discourse uses and temporal uses remains unclear. Subsequent contributions to the discourse necessarily take place in the future, and previous ones in the past. Whilst for Lakoff discourse and temporal uses are classified as different functions of demonstratives, it seems that the borderline between the categories is blurred in many cases. Lakoff also points out some subtle distinctions and nuances

² This utterance is perfectly acceptable if interpreted in a gestural sense: for example, whilst pointing at a written list of courses which the speaker has already taken.

associated with discourse uses of demonstratives. She suggests that whilst ‘that’ produces ‘a more colloquial tone’ in some examples, such as (18), it results in a distancing effect in others, such as (19) (Lakoff 1974, pp. 349-350).

(18) That man’s gonna get his one of these days!

(19) Kissinger made his long-awaited announcement yesterday. That statement confirmed the speculations of many observers.

Diessel distinguishes two distinct categories for demonstratives that refer to elements within the discourse: anaphoric and discourse deictic. In his view ‘anaphoric demonstratives are coreferential with a noun or noun phrase in the previous discourse’ (1999, p. 95), as in (6) (‘The cowboy entered. This man was not someone to be messed with’) above. By contrast, discourse demonstratives refer to propositions, focusing ‘the hearer’s attention on aspects of meaning expressed by a clause, a sentence, a paragraph or an entire story’ (1999, p. 101), as in (20):

(20) A: I’ve heard you will move to Hawaii?

(21) B: Who told you that?

5.2.1.3 Symbolic and emotional uses

Fillmore and Lakoff diverge considerably more in their description of the third and final category of demonstrative use. For Fillmore, this covers ‘symbolic’ uses, which he defines as those requiring the hearer to ‘know...certain aspects of the speech communication situation’ in order to understand the utterance correctly. He gives the example of a lecturer at a university using the phrase ‘this campus’ as part of an utterance (1997, p. 63). The lecturer provides no explicit demonstration of the referent in this case, and so the use cannot be termed gestural. However, the meaning is taken to be ‘the campus in which I am now located’, and Fillmore classes this use of the demonstrative as ‘symbolic’ of the speaker’s meaning. Diessel (1999) and Levinson (2004) describe these uses as ‘symbolic exophoric’.

Lakoff’s final category contains what she terms ‘emotional’ uses. In these cases, the use of a demonstrative, and the choice between distal and proximal forms, creates effects that go beyond simply securing reference. Lakoff divides this category into

three subcategories. The first contains uses of 'this' which are associated with a 'camaraderie' or 'closeness' effect, or are used to create a sense of 'vividness' (Lakoff 1974, p. 347). Consider (22):

(22) A: I see there's going to be peace in the mideast.

B: This Henry Kissinger really is something!

According to Lakoff, such cases occur when 'the speaker wishes to allude to something, or someone, already mentioned, but outside the discourse proper...its most natural use seems to be with proper names that the speaker expects the hearer to be familiar with' (Lakoff 1974, p. 347).

Lakoff's second category of 'emotional' demonstratives involves cases where 'this' seems to stand in the place of the indefinite article 'a', as in (23).

(23) There was this travelling salesman, and he...

She describes this as a colloquial use which produces a 'vividness' effect. It has been analysed more recently as a case of 'specific indefinite' use, in which the speaker indicates that she has a specific individual in mind and expects the hearer to treat that individual as a discourse referent (Maclaren, 1980).

Finally, Lakoff presents examples which combine emotional and discourse deixis. In these cases, 'a person has been referred to in one sentence, and the next provides additional information about him pertinent to the subject-matter of the prior sentence'. The examples she gives from this category all use the demonstrative in its bare form, for instance (24).

(24) 'Don't lie to me,' said Dick. This was a man who had twice been convicted of perjury.

Again, she suggests that a sense of closeness is associated with use of the proximal demonstrative, and that this is understood as reflecting the closeness or relevance of the second sentence to the first.

According to Lakoff, whereas 'spatio-temporal' uses of 'this' and 'that' have largely opposite effects, in 'emotional' uses the opposition is somewhat neutralised.

Uses of ‘this’ to convey camaraderie are paralleled by uses of ‘that’ which appear to create a sense of intimacy. Consider (25) and (26) as spoken by a doctor to a patient.

(25) How’s that throat?

(26) How’s your throat?

Lakoff suggests that despite the use of the distal demonstrative, (25) conveys a sense of closeness or intimacy, and that had the doctor chosen to utter (26) instead, she would have been drawing attention to the fact that the ailment is not hers, and thus creating a distance between herself and her patient. According to Lakoff ‘that’ may also be used to evoke ‘emotional solidarity’ (p.352) between speaker and hearer by implying that they both share the same view of the discourse subject matter. As she points out, this unifying effect arising from the use of a demonstrative usually associated with distance is perhaps unexpected.

These ‘emotional’ uses are categorised as ‘recognitional’ by Diessel (1999), who treats them as signalling that information is ‘discourse new’, ‘hearer old’ and ‘private’. ‘Private’ information is ‘information that speaker and hearer share due to common experience in the past’, as opposed to ‘general cultural information’ that is shared by all members of the speech community (Diessel, 1999, p. 106). According to Diessel, this use of demonstratives suggests that the speaker and hearer share the same viewpoint, and therefore indicates ‘emotional closeness, sympathy and shared beliefs’ (p. 107).

Cornish (2001), following Strauss (1993), Cheshire (1996) and Glover (2000), also treats some ‘emotional’ (or ‘empathetic’, as he and Diessel term them) uses of ‘that’, such as (27), as creating a ‘solidarity’ effect:

(27) ...‘Do we want machines which are more intelligent than humans, or should we call a halt to it?’ he asked. ‘We are still a long way from *that* decision but I think...’ (The Guardian) (Cornish, 2001, p. 303)

According to Cornish, ‘the speaker is tacitly instructing the addressee to place the referent outside his/her discourse-cognitive sphere’, and thus casting the hearer as ‘a potential ally in the speaker/writer’s argumentative stance’ (Cornish, 2001, p. 304).

At this point in his discussion, Cornish suggests that we view these interpretive effects as ‘a type of inference which may be drawn from the use of ‘that’ in context,

rather than it being a basic motivating principle determining its use' (Cornish, 2001, p. 305). So we move away from simply identifying various uses of 'this' and 'that', and towards an account where an underlying meaning interacts with the discourse context to yield more and different effects. This move will prove to be pivotal in the development of my relevance-based analysis.

5.2.1.4 Towards a unified approach

Lakoff's and Fillmore's descriptions of the effects that result from the choice of demonstrative seem, for the most part, intuitively correct. However, neither gives a systematic explanation of why these effects should arise, and how they fit into an overall account of utterance interpretation. Both seem to be more concerned with providing a descriptive taxonomy rather than considering the semantic and pragmatic functions of demonstrative determiners.

Furthermore, as mentioned above, the boundaries between the categories they propose seem at times blurred and arbitrary. It is unclear how Fillmore's temporal cases fit into the taxonomy, and the generalisations made appear to have many exceptions. Lakoff makes the interesting suggestion that in some cases the contribution of the demonstrative goes beyond merely securing reference. However, her category of 'emotional' uses amounts to little more than a bundle of disparate examples. In analysing these uses and their effects, she appeals to unanalysed notions of 'vividness', 'closeness' and 'camaraderie' which are standard in rhetorical accounts of stylistic and poetic effects. Relevance theory has tried to go beyond such vague notions, and aims to offer an explanation based on the relation between indirectness, effort and effect (Sperber & Wilson, 1986/95, pp. 217-224; Blakemore, 1993; 1994). I will return to these examples in developing my own analysis of demonstratives, and consider how relevance theory might approach them.

Both Diessel (1999) and Cornish (2001) provide accounts which seek to link the various uses of demonstratives. Diessel argues that the exophoric use is 'basic and unmarked' (p. 114), and that the other uses are derived from it. By contrast, for Cornish, it is the 'modal' or 'empathetic' use that is basic, and the other uses which are derived (Cornish, 2001, p. 312). The central claim of my account will be that demonstratives encode procedures, and that these procedures interact with the discourse context to yield the various uses and interpretations. Like Diessel and Cornish, I assume that we must move beyond the traditional characterisation of the difference between the distal demonstrative 'that' and its proximal counterpart 'this' as relating

only to the relative spatial distance between the speaker and the entity to which she is referring (Lyons, 1977; Levinson, 1983; Anderson & Keenan, 1985; Halliday, 1985). Much evidence suggests that this ‘static’ characterisation is inadequate (Strauss, 1993; Janssen, 1995; Maes, 1996; Cornish, 2001; Strauss, 2002). Both Lakoff and Fillmore treat ‘distance’ as going beyond the purely spatial dimension to encompass cases of symbolic, metaphorical and emotional distance, although neither offers a truly unified account of the data. Indeed, Lakoff is fairly pessimistic about the prospect of providing a unified account, and comments that whilst the task of uniting spatial, discourse and emotional uses is the ultimate aim, ‘How this is to be done remains mysterious’ (Lakoff 1974, p. 355).

However, there have been various attempts to bring demonstratives within the scope of more general analyses of referring expressions such as those proposed by Ariel and Gundel, which treat different referring expressions as signalling different types of cognitive status. As discussed in chapter 3, Ariel’s Accessibility Theory (1990) treats demonstrative pronouns as marking intermediate accessibility, and claims, more specifically, that ‘more accessible entities are referred to by the proximal marker. Less accessible ones will be referred to by the distal ones’ (1990, p. 51). She endorses Fillmore’s suggestion that “‘that’ requires identifiability by both speaker and addressee, whereas ‘this’ sometimes refers to objects accessible only to the speaker’ (1990, p.53). And, like Fillmore and Lakoff, she also discusses the subtler effects that choice of a demonstrative determiner can have on utterance interpretation. Consider the examples in (28) and (29).

(28) That holiday we spent in Cyprus was really something.

(29) The holiday we spent in Cyprus was really something.

According to Ariel, use of the demonstrative rather than the definite description in (28) has the effect of raising ‘into consciousness “livelier pictures” of the said holiday’ (1990, p. 54), presumably because of the higher level of Accessibility signalled by this form.

Gundel, Hedberg and Zacharski (1993) offer a slightly different analysis of demonstrative determiners. According to the Givenness hierarchy, use of a definite article signals that the intended referent is ‘uniquely identifiable’, use of a distal demonstrative signals that the referent is ‘familiar’, and use of a proximal determiner

signals that the referent has the even higher cognitive status of being ‘activated’³. Thus, the choice of determiner is seen as guiding the hearer to the speaker’s intended referent.

Along similar lines, Strauss (2002) analyses the different forms of demonstrative as signalling different levels of ‘focus’, where focus is defined as ‘the force with which the hearer is instructed to seek the referent’ (Strauss 2002, p. 135). Use of ‘this’ signals that the speaker regards the information as important, and as therefore deserving a high degree of focus. Use of ‘that’ signals ‘medium focus’, and ‘it’ signals low focus. On this account, a speaker’s choice of referring expression is based on ‘spontaneous, contextually grounded interactional factors between the various participants involved in the talk’ (Strauss, 2002, p. 133), including the relationship between speaker and hearer, the amount of information they share, and the relative importance of the referent to the speaker.

Although these three approaches differ in their details, all three seek to place their account of demonstratives within a cognitively-orientated analysis of referring expressions in general, and move away from physical notions such as the referent’s proximity to or distance from the speaker. They share the idea that the hearer is directed to a representation of the intended referent from amongst the set of candidate referents via information encoded by the various referring expression forms.

Continuing this theme, Reboul (1997) situates her analysis more directly within the relevance-theoretic framework, and introduces the idea that demonstratives might encode non-truth conditional meaning that influences interpretation. Unlike Ariel, Gundel and Strauss, however, she suggests that the difference between ‘this’ and ‘that’ has ‘to do with the relative distance between the speaker and the object designated’ (Reboul 1997, p. 11), thus reintroducing the traditional appeal to proximity or distance. However, she remains rather vague about how this difference might be developed within a full account. In the next section, I take Reboul’s proposal that demonstratives encode non-truth conditional meaning as a starting point, although like Strauss and Cornish, I will move away from the idea that demonstratives encode purely physical notions of proximity and distance relating to the ‘geographical location of their referents in relation to the speaker’ (Cornish, 2001, p. 306), and go for a more abstract proximity-based account. My main aim is to show how demonstratives can be analysed in terms of encoded procedures. By applying these procedures to the data, I will try to provide evidence that the uses of demonstratives fall into two broad categories which cross-cut

³ I refer here to the complex demonstrative forms ‘This N’ or ‘That N’. According to the Givenness hierarchy, bare demonstratives, both the distal ‘that’ and proximal ‘this’, signal that the intended referent is ‘activated’. See Gundel, Hedberg and Zacharski (1993) for further discussion.

the tripartite distinctions proposed by Lakoff and Fillmore, and unify the various uses identified by Diessel and Cornish.

First, I will consider some cases where the choice of a distal as opposed to a proximal form changes the way that reference is likely to be resolved. I will then examine some examples where it does not. When the choice of determiner does not affect reference resolution (and therefore does not affect the derivation of explicit content), we must ask what role the encoded procedure might play. I will argue that the range of different uses discussed above are, in fact, a result of the same underlying procedure interacting with different discourse contexts to yield a wide range of inferential effects. This is a development of Diessel's and Cornish's ideas, and Lakoff (p.346) herself makes a suggestion that seems to be moving in this direction:

we may want to consider this [spatio-temporal] use as exemplifying the 'basic' demonstrative, with the others perhaps to be derived from it by a process of metaphorization, or abstraction (Lakoff, 1974, p. 346).

The cognitively-orientated accounts of Ariel, Gundel, Hedberg and Zacharski, Strauss and Reboul take us closer to a unified analysis, but still fail to address the full range of data. In the following sections, I will try to develop such a unified analysis using the relevance-theoretic notion of procedural meaning, and show how the inferential effects of the various uses of demonstratives might be derived via interaction between procedural meaning, discourse context and the relevance comprehension procedure.

Before embarking on this, however, I will look briefly at some observations by the philosopher David Kaplan (1989) on the treatment of demonstratives and their role in reference resolution, which raise important questions about the value of cognitively- and communicatively-oriented analyses of the type I have been considering here.

5.2.1.5 Kaplan: Pure indexicals, true demonstratives and intentions

Kaplan (1989) distinguishes between what he terms 'pure indexicals' and 'true demonstratives'. A true demonstrative requires some sort of accompanying demonstration to make reference. For pure indexicals, on the other hand, no demonstration is required to establish reference, and 'any demonstration supplied is either for emphasis or is irrelevant' (p.491). Some expressions can function both as pure indexicals and as demonstratives, and Kaplan provides the examples in (30) and (31) to illustrate.

(30) I am here.

(31) In two weeks, I will be here. [pointing at a city on a map]

For Kaplan, 'here' in (30) is a pure indexical, whereas in utterance (31) it functions as a demonstrative, and reference will only be resolved if there is an accompanying gesture or other clue to the speaker's intentions. This aspect of his theory is summarised in the principle in (32).

(32) The referent of a pure indexical depends on the context, and the referent of a demonstrative depends on the associated demonstration.

I will return to this distinction between indexicals and true demonstratives when I consider the uses of demonstrative determiners and their contribution to explicitly communicated meaning.

In *Afterthoughts* (1989), Kaplan revisits these ideas and discusses true demonstratives in more detail. Although he repeats that demonstratives must be accompanied by a demonstration of some kind, he argues that what determines reference in these cases is the speaker's intention rather than the external gesture or indication. The demonstration is there, he claims, 'only to help convey an intention' (pp. 583-584), and we should 'regard the demonstration as a mere externalization of this inner intention,' which itself determines the referent (p. 582). This shift in emphasis from an external demonstration to the intention of the speaker brings us closer to pragmatically-orientated approaches such as relevance theory.

In relevance theory, speakers' intentions are seen as playing a much greater role in reference resolution than is envisaged in many formal semantic accounts. As a result, there is no need for a theoretical distinction between 'true demonstratives' and 'pure indexical' uses of demonstratives. Instead, all referential uses of 'this' and 'that' are seen as dependent on the speaker's intentions, whether or not they are accompanied by extra, external clues. Indeed, in a footnote, Kaplan himself states that he views 'demonstrations as playing the same role for true demonstratives as does pointing at oneself when using the first-person pronoun' (n35, p.582). Since the first-person pronoun is one of the clearest examples of a pure indexical in Kaplan's sense, this move towards recognising the role of intentions blurs the boundaries of Kaplan's original

distinction, and brings us closer to an analysis where all referential uses ultimately depend on the speaker's intentions.

Bach (1992), who defends and discusses Kaplan's view on intentions, argues explicitly that the referential intention is part of the wider communicative intention: it 'isn't just any intention to refer to something one has in mind but is the intention that one's audience identify, and take themselves to be intended to identify, a certain item as the referent by means of thinking of it a certain identifiable way' (Bach, 1992, p. 143). Moreover, the communication of this intention does not happen 'by magic', and having the intention in one's mind does not guarantee that it will be successfully recognised. Rather, 'you decide to refer to something and try to select an expression whose utterance will enable your audience, under the circumstances, to identify that object' (1992, p. 145). Thus, the recognition of intentions is all-important to reference resolution, and the speaker should give whichever linguistic or non-linguistic clues are needed to secure it. To consider in more detail how this plays out in terms of utterance interpretation, I return again to the framework of relevance theory, and in particular, to the notion of procedural meaning.

5.2.2 A procedural analysis of demonstratives

Since the idea of procedural meaning was first developed by Diane Blakemore (1987) there have been various attempts to characterise the individual procedures encoded by different forms of expression (Blakemore, 2000, 2002; Wharton, 2003; Iten, 2005; Hall, 2007; Hedley, 2007). As Wilson and Sperber (1993) point out, this is no easy matter, since 'We have direct access neither to grammatical computations nor to the inferential computations used in comprehension' (1993, p. 16). As a result, procedural meaning is very hard to pin down. Still, it is now quite widely assumed in relevance theory that personal pronouns can be analysed in procedural terms, so that the pronoun 'she', for instance, facilitates the retrieval of female (or grammatically feminine) candidate referents (see Wilson and Wharton 2006). In the following sections, I will argue that demonstratives should also be analysed as encoding procedural information, and consider what this procedural information might look like.

According to relevance theory, procedural information guides the inferential phase of comprehension, and inference may play a role on both sides of the explicit/implicit divide. Assuming this model, I consider how the same underlying procedure might play a role not only in reference resolution, but also in the derivation of implicatures. This should bring us closer to a unified account of the various functions

of demonstrative forms. Bearing in mind the wide range of uses identified by Fillmore, Lakoff and Levinson, I begin by considering in more detail how traditional notions such ‘proximal’ and ‘distal’ might fit into the framework, and how a speaker’s conception of relative proximity or distance might influence their choice of demonstrative. In this, I depart from Ariel’s and Gundel’s accounts, in which such traditional notions appear to play no role.

Reboul (1997) argues that there is a clear difference in acceptability between examples (33) and (34), although they refer to the same plane, express the same propositions, and should therefore share their truth conditions.

(33) A plane crashed yesterday in New York. This plane flew every day from Miami to New York.

(34) ?A plane crashed yesterday in New York. That plane flew every day from Miami to New York.

Since the two examples differ only in the choice of demonstrative determiner, ‘this’ or ‘that’, this difference in acceptability must be linked to the choice of determiner. And since the same proposition, with the same truth conditions, is expressed in both cases, the difference does not seem to be a conceptual one, and we are left with the possibility that the demonstrative determiners ‘this’ and ‘that’ differ in the procedural information they encode.

What might this procedural information look like, and how might a speaker use it to optimise the relevance of her utterance? I take as my starting point the traditional distal/proximal distinction often used to characterise the difference between ‘this’ and ‘that’, and explore how we might generalise the notions of proximity and distance to apply to a much wider range of cases.

Powell (1998, p. 18), who proposes a procedural approach based on the traditional distal/proximal distinction, suggests that ‘this’ and ‘that’ encode related, but different, two-step procedures. ‘This’ encodes the procedure, ‘find the speaker and then find an object near the speaker,’ and for ‘that’, the procedure is the same except that the hearer expected to find an object far away from the speaker. This analysis is then extended to the plural forms, with ‘these’ encoding the procedure, ‘find an object near the speaker and then find a plural referent relating to it’, and *mutatis mutandis* for ‘those’. Powell’s procedures are thus built around the notions of proximity to and of distance from the

speaker. This approach highlights the important point that proximity/distance are relational terms. A candidate referent will be proximal or distal relative to some point of reference. Here, proximity and distance are not properties inherent in the referent itself, but depend on the context in which the utterance is processed, and will vary as elements of the discourse context vary. This is particularly evident in an exchange such as (35)-(37).

(35) Natasha : Pass me that book.

(36) Marya : This book?

(37) Natasha : Yes, that book.

In this case, each speaker assesses the proximity of the book relative to herself, and from her own point of view, so that the same entity is referred to using both the distal and proximal forms. Once the book has been passed to Natasha, she may then utter (38) and still be understood as referring to the same item.

(38) Natasha : I love this book.

From a relevance-theoretic perspective, drawing attention to the relative proximity of an entity to a speaker will be relevant if it reduces the effort involved in reference resolution, either by adding an extra layer of activation to the hearer's mental representation of the intended referent, or by ruling out any non-intended competitors. If the relative proximity to the speaker of the intended referent sets it apart from the other potential referents in either of these ways, then providing such proximity information is likely to be an efficient strategy for the speaker to follow.

A spatially-based notion of proximity where 'this' is used to refer to objects close to the speaker and 'that' is used to refer to objects far from the speaker may seem intuitively attractive in these examples. However, it has been argued using evidence from psychology that this approach is over-simplistic and problematic. Kemmerer (1999) presents evidence that our perceptual assessments of proximity to and distance from ourselves do not fit well with the way that distal and proximal demonstratives are used cross-linguistically. Perceptually, a basic distinction is drawn between objects that are roughly within arm's length and those that are beyond arm's length. However, the

use of proximal and distal demonstratives does not correspond to this distinction. Rather, according to Kemmerer, the spatial distinctions conveyed by the use of demonstratives, are much more abstract and relative.

Consider examples (39) and (40) (Kemmerer 1999, p. 52, taken from (Talmy, 1988)).

(39) This speck is smaller than that speck.

(40) This planet is smaller than that planet.

The specks referred to in (39) may be tiny, close together and within the speaker's reach, whilst the planets in (40) will be huge, far apart and remote from the speaker. In both cases, however, 'this' and 'that' may be used felicitously to refer to the individual specks and planets, respectively. Just as 'smaller' is a relative notion dependent on the discourse context and the nature of the items it describes, so the proximity information conveyed by the demonstratives is also relative. According to Kemmerer, 'the proximal demonstrative *this* means simply "closer to the deictic center than *that*" and, conversely, the distal demonstrative *that* means simply "further from the deictic center than *this*' (1999, p. 52). So a story begins to emerge where 'this' and 'that' are to be interpreted not only relative to a context and a deictic centre (e.g. a speaker), but also relative to one another. This suggests that 'this' and 'that' are fundamentally contrastive in nature. Use of a demonstrative may be seen as adding an extra layer of activation to the hearer's representation of the intended referent by setting it in contrast with similar, but different, competing representation(s).

Kemmerer sums up by saying that demonstratives involve 'semantic representations rather than concrete spatial representations; and the remarkable pragmatic flexibility of demonstratives is due to the fact that they are essentially deictic terms that cannot function apart from specific discourse contexts' (1999, p. 56). This notion of relativity and the importance of the role played by discourse context will be a central theme in my relevance-based account.

As mentioned in passing above, a further objection to analysing demonstratives in terms of purely spatial notions of proximity is that the contrasts they evoke are not restricted only to the spatial domain. As Lakoff and Fillmore point out, an entity may be distinguishable by its relative proximity to the speaker in place, time, thought or discourse, and assessments based on these dimensions may not always coincide. For

instance, a referent may be seen as proximal in discourse terms if it is the current focus of the exchange, although it may be distal in spatial terms. Take, for example, a case of so-called deferred reference, as discussed by Nunberg (1995; 2004). A customer hands a car key to a car park attendant and utters (41).

(41) This is parked out back.

Here, the speaker uses the demonstrated object (the key) as a proximal clue to the intended referent (the car), which itself is distal. The hearer's recognition of the speaker's intention to refer to the car (via the clue provided by the key) brings the mental representation of the intended referent into the current focus of the exchange. A representation of the intended referent (the car) is thus proximal in thought and discourse, although the object itself remains spatially distant.

Thus, assigning reference to a demonstrative expression requires identification of the deictic centre, on the one hand, and of the dimension in which a contrast is to be drawn, on the other. Identification will proceed, as always, via the relevance-theoretic comprehension procedure. We therefore predict that a speaker aiming at optimal relevance will construct any demonstrative expression so that the hearer can test interpretative hypotheses in order of accessibility, and stop when his expectations are satisfied. Although the deictic centre will typically be the present time, location or status/role of the speaker, this is not necessarily the case. For example, consider the exchange in (42)-(43).

(42) Dentist: Does this one hurt?

(43) Patient: Yes, it's that one.

Fillmore (1997, p.123) notes that the speaker may use a distal form for something relatively close to her in order to present it from the hearer's perspective. As the tooth in question is in the patient's mouth, a speaker-based account of proximity relations would predict that the dentist should use the distal form and the patient the proximal form. However, in this discourse context, it is the viewpoint of the dentist, who has presumably used a gesture of some sort to identify the intended tooth, that is relevant, and so both interlocutors represent the situation from the dentist's point of view, and treat him as the deictic centre.

My argument so far might be summarized as follows. Reference is resolved via recognition of the speaker's intentions. Demonstratives are used when the nominal part of the referring expression (if there is one) is not uniquely denoting in the discourse context. The speaker may then exploit procedural information encoded by the demonstrative to guide the hearer to the intended referent (i.e. indicating to which instance of 'N' she intends to refer)⁴. This procedural information relates to the proximity/distance of the intended N to a deictic centre, as compared to competing instances of Ns. The effect should be to single out the intended referent and add an extra layer of activation to its representation so that it will be the most accessible one in the discourse context. 'This' is used when the intended referent is 'closer to the deictic center than "that"' (where 'that' would be used to refer to a competing instance of N). 'That' is used when the intended referent is 'further from the deictic center than "this"' (where 'this' would be used to refer to a competing instance of N). Proximity and distance may be assessed on various dimensions, e.g. in spatial, temporal, emotional, thought or discourse terms. The hearer determines the deictic centre and the appropriate dimension of assessment in the same way as with other pragmatic processes: by following the path of least effort and stopping when the expectations of relevance raised by the utterance are satisfied.

Although in some cases the choice of a demonstrative determiner directly affects reference resolution, in other cases it appears to contribute instead to the implicitly communicated meaning. In the remainder of this section, I consider examples of both, and try to show how they are all analysable in terms of the procedure encoded by demonstratives and the overall aim of satisfying expectations of optimal relevance.

5.2.3 Demonstrative determiners and explicatures

5.2.3.1 'This' and 'that'.

According to relevance theory, 'an assumption communicated by an utterance U is explicit [hence an "explicature"] if and only if it is a development of a logical form encoded by U' (Carston, 2004, p. 635). In this section I consider cases where choice of demonstrative determiner affects the reference resolution process. As reference assignment involves a development of the encoded logical form, the choice between

⁴ She may also use conceptual information to do this – e.g. uttering 'the black cat' rather than 'the cat' so that the nominal is no longer uniquely denoting.

'this' and 'that' directly affects what is explicitly communicated in such cases. Consider the examples in (44)-(47).

(44) I'll have a slice of this cake.

(45) I'll have a slice of that cake.

(46) We went for a lovely walk this afternoon.

(47) We went for a lovely walk that afternoon.

If (44) and (45) are uttered in the same discourse context where the speaker has been offered the choice of two cakes, we can imagine that reference might well be resolved differently in each case. Similarly, the resolution of the complex demonstratives in (46) and (47) will be different. In both pairs, the only difference is the form of the determiner, and in both cases, substitution of the complex demonstrative with a corresponding definite description will be infelicitous, as in (48) and (49).

(48) ? I'll have a slice of the cake.

(49) *We went for a lovely walk the afternoon.

Although (48) is not ungrammatical in itself, it will be infelicitous in a discourse context where there are two cakes whose representations are equally accessible, since reference will not easily be resolved. However, whilst in examples (44)-(47) the demonstratives clearly contribute to the explicitly communicated meaning, they do not necessarily function in exactly the same way. Utterances (44) and (45) are only likely to be felicitous if accompanied by a gesture indicating the intended referent. This is not the case for (46) and (47). These judgements recall Kaplan's distinction between pure indexicals and true demonstratives, as outlined in section 5.2.1.5. Taking this distinction as a starting point, I consider how the interpretations and judgements in (44)-(49) might be accounted for.⁵

⁵ Following the discussion in section 5.2.1.5, I am not committing to the view that there are pure indexicals which are assigned reference independently of the speaker's intention. However, there are some cases where the hearer may not need any additional clues beyond the information encoded in the demonstrative in order to recognize that intention. I retain Kaplan's terminology for ease of exposition.

5.2.3.2 Demonstrative uses

I begin by considering examples where the use of the complex demonstrative is what Kaplan would consider to be truly demonstrative. Consider again examples (44) ('I'll have a slice of this cake') and (45) ('I'll have a slice of that cake') in a discourse context where there are two cakes. Here, representations of both cakes may be activated and highly accessible, and yet the hearer must determine which of them the speaker intends to pick out. In such a case, 'this' and 'that' function as true demonstratives. They are only appropriate when the intended referent is in the physical context, and their utterance is only likely to be felicitous if there is some other clue to the speaker's intention, such as a physical demonstration of where the speaker's attention is focused. Whilst such a demonstration may be as explicit as a pointing gesture, it may also be a subtle cue such as eye gaze direction or directional nodding. As Kaplan notes, the demonstration itself is 'a mere externalization of the perceptual intention' (1989, p. 583), and it is this intention that determines the referent. In these cases, however, something beyond the linguistically encoded meaning is required to convey this intention.

The contrast evoked by the use of a proximal/distal form as opposed to a definite description sets the intended referent apart from competing referents. Whilst the extra layer of activation is added to its representation as a result of the speaker's indication of where her attention is focused, use of a demonstrative form encourages the hearer to take this attention into consideration. As Neale (2007, p. 103) suggests, using a demonstrative form to refer to an item ϕ draws 'attention to the fact that not any old ϕ will do.' By drawing attention to a particular entity in the physical environment, the speaker increases the accessibility of its representation, so that the hearer will test it first as a potential referent. The speaker's choice between 'this' and 'that' again comes down to the underlying procedures encoded by the terms, and the clues they give the hearer as to how inferential comprehension should proceed.

Taking the speaker as the deictic centre, use of 'this', in this type of example, will encourage the hearer to expect a referent that is, in some salient way, relatively proximal to the speaker compared to another competing candidate referent. Accordingly, a hearer would be confused if the speaker were to utter (28) ('I'll have a slice of this cake') whilst indicating a cake that was further away from her than some other salient cake. Thus, in those examples that Kaplan would call truly demonstrative, the proximity information encoded by the demonstrative determiner is a clue to where

the speaker's attention is likely to be focused, and therefore what her referential intention is likely to be. However, in these cases, as Kaplan points out, extra-linguistic clues are also needed.

5.2.3.3 Indexical uses

The contribution of demonstratives to reference resolution, and hence to the explicitly communicated meaning, is not limited to Kaplan's true demonstrative uses of 'this' and 'that'. Examples such as (46) and (47), or (50) and (51), are perfectly felicitous without any accompanying gesture or further clue to the speaker's intention. Under Kaplan's original definition, they are pure indexicals.

(46) We went for a lovely walk this afternoon.

(47) We went for a lovely walk that afternoon.

(50) In this section we will continue the argument.

(51) In that section we will continue the argument.

In examples of this type, there is invariably more than one potential referent, since neither 'afternoon' nor 'section' is uniquely denoting. However, in both cases involving the use of 'this', there is only one instance of each type that is likely to be considered proximal relative to the deictic centre of the speaker at the time of utterance. In both cases, too, the relevant type of proximity is likely to be temporal or discourse-related rather than spatial; 'this afternoon', will generally refer to the afternoon of the current day, and 'this section' will generally refer to the section in which the current utterance occurs⁶. Utterances of this type can occur discourse initially, since there is no need for any further contextual information.

The situation is slightly different for indexical uses of the distal demonstrative 'that', as in (47) and (51). Whereas there is only one afternoon or one section which is likely to count as proximal, there will be more than one which could count as distal. 'That afternoon' could refer to any afternoon on another day than the day of utterance, and 'that section' could refer to any section apart from the current one. Following the

⁶ There are corresponding true demonstrative readings available for these utterances. If, whilst poring over a holiday diary, a speaker points at a particular date and utters (46), the speaker's manifest intention to refer to the afternoon of the demonstrated day would override the indexical interpretation.

analysis outlined in section 5.2.2, I predict that ‘that’ will be used when the intended referent is further from the deictic centre than another candidate which could have been picked out by use of ‘this’. As we have seen, ‘this’ picks out the current day or section, so ‘that’ could be used to refer to any other day or section. However, interpretation proceeds in the usual way with the set of potential referents being narrowed to include only those which are relatively further away from the deictic centre than the current instance⁷. In effect, this rules out the afternoon of the present day, and the current section, respectively. The hearer then follows the path of least effort, testing potential interpretations in order of accessibility. Thus, for an utterance of (47) (‘We went for a long walk that afternoon’) to be felicitous, there should be one non-proximal afternoon whose representation is more accessible than those of the others⁸.

Consider (52) and (53).

(52) We are going on holiday this week.

(53) We are going on holiday that week.

Utterance (52) should be felicitous in any discourse context. The use of ‘this’ constrains the set of potential referents to just those which are proximal to the deictic centre relative to competing candidate referents. Other things being equal, the only member of this set will be the week of the utterance. By contrast, (53) will only be felicitous when there is a particular non-proximal week whose representation is more accessible than those of any others. Use of the distal demonstrative in effect rules out the current week, but in doing so, it still leaves all other weeks as possible referents. For (53) to be felicitous, a representation of one of these candidate referents must be more accessible than those of the others. Utterance (53) might, for example, be appropriate if the speaker and hearer have been discussing when they might meet up next, and the hearer has suggested a particular date. In that case, a representation of the week which includes that date will be more accessible than those for other non-proximal weeks. Whilst acceptable in such a discourse context, (53) would be very strange as a discourse initial utterance.

⁷ Alternatively, we might view this as adding an extra layer of activation to the representations of the non-current instances. However, the result is the same and it’s not clear how the two possibilities might be empirically distinguished.

⁸ Or a highly accessible contextual assumption that combines with one potential referent more easily than the others to result in an interpretation that is relevant in the expected way.

This pattern in the use of demonstratives has been acknowledged in the various different analyses of ‘that’, and strikingly similar conclusions have been drawn about when it may be used felicitously. According to Fillmore (1997, p. 105), ‘that’ is used when both interlocutors are aware of what is being talked about. For Ariel (1990, p. 53) ‘that’ requires identifiability by both speaker and addressee, and Gundel, Hedberg and Zacharski (1993) capture a similar intuition by associating ‘that N’ with the cognitive status ‘familiar’. Such generalizations fall out naturally from a relevance theory approach where interpretations are tested in order of accessibility. On such an approach, what makes the use of ‘that’ felicitous in these discourse contexts is not the fact that both interlocutors necessarily know what is being talked about. Rather, the felicitous use of ‘that’ depends on one representation’s being more accessible to the hearer in the discourse context than its competitors. Discourse contexts of this type are highly likely to coincide with those where the intended referent fits the theory-specific definitions of ‘familiar’, ‘identifiable’ and so on.

In sum, and having considered both ‘true demonstrative’ and ‘pure indexical’ uses of complex demonstratives, it seems that both categories may contribute to the explicit content of an utterance. Kaplan’s distinction boils down to a difference in the nature of the clues that the speaker provides to help the hearer retrieve the explicit content. In the pure indexical cases, the nature of the intended referent and the discourse context are such that the linguistically encoded meaning is enough. In the true demonstrative cases, however, the linguistic meaning is not sufficient, and further non-linguistic clues are needed. In a relevance-based account, however, these are just different means to the same end. The speaker intends the hearer to pick out a certain referent, and provides whatever linguistic or non-linguistic clues are necessary to achieve this aim.

5.2.4 Demonstrative Determiners and Implicatures

5.2.4.1 Explicit or implicit: a test

In the previous section, I argued that choice of demonstrative determiner can affect what is explicitly communicated by an utterance. In the cases considered, the choice between ‘this’ and ‘that’, or between a complex demonstrative and a definite description, affects the proposition the speaker is taken to have expressed. However, there are other utterances where this does not seem to be the case. In such examples, substituting a complex demonstrative form for a definite description, or a proximal for a distal

demonstrative, or vice versa, does not affect the way that reference is resolved.

Consider examples (54)-(55) and (56)-(59).

(54) Tartan was very popular in the 19th century. Prince Albert helped this trend by developing his own Balmoral tartan.

(55) Tartan was very popular in the 19th century. Prince Albert helped the trend by developing his own Balmoral tartan.

(56) This section focuses on what that procedural information might look like.

(57) This section focuses on what this procedural information might look like.

(58) This section focuses on what the procedural information might look like.

(59) This section focuses on what such procedural information might look like.

The use of ‘this’ rather than ‘the’ in (54) does not appear to affect the explicit content of the utterance. Version (55) is not only also acceptable, but reference is resolved in the same way as in (54). This case contrasts with those in (44)-(45) above (I’ll have a slice of this/that cake), where replacing the demonstrative with ‘the’ in a multi-cake discourse context leads to infelicity.

It seems, then, that replacing a demonstrative determiner with the definite article might provide a test for whether or not the demonstrative adds anything to the explicit content of an utterance that could not equally well have been conveyed by ‘the’.

Consider utterance (60), taken from Powell (2001).

(60) That dog with three legs is called ‘Lucky’.

As Powell points out, the role played by the complex demonstrative ‘that dog’ in (60) depends on the context in which it is uttered. If (60) is uttered whilst ‘standing in a room full of dogs, all but one of which have the standard canine allocation of legs’ (Powell 2001, p. 62), then the demonstrative determiner can be replaced by the definite article without affecting the felicity of the utterance, as in (61).

(61) The dog with three legs is called 'Lucky'.

However, if the speaker is standing in a room full of three-legged dogs, then the substitution leads to infelicity and probable reference failure. In the first discourse context, the conceptual information encoded by the nominal is sufficient to narrow the set of potential referents to a point where a representation of the intended referent (i.e. the dog with three legs) is the most accessible one. In the second discourse context, where there are several three-legged dogs, this is not the case, and further clues, in the form of information about relative proximity, probably combined with some type of non-linguistic demonstration, will be required⁹.

Examples (56)-(59) show that in certain discourse contexts the speaker may have an even wider range of options available when constructing her referring expression. Again, in these cases, the substitution of one determiner for another does not seem to affect the way that reference is resolved, suggesting that the encoded conceptual information alone is enough to guide the hearer to the explicit content of the utterance.

5.2.4.2 Demonstratives and inferential effects

Given the data in (54)-(59), how can we account for the various acceptable forms, and what effect might the choice of determiner ultimately have on interpretation in such cases? To address these questions, I once again return to the relevance theoretic comprehension procedure, as discussed in chapter 4.

A hearer following the relevance-theoretic comprehension procedure should test interpretations in order of accessibility, and stop when his expectation of (optimal) relevance is satisfied: that is, when he has enough cognitive effects to make the utterance at least relevant enough to be worth processing, and the most relevant one compatible with the speaker's abilities and preferences. Although the interpretation process involves identification of explicit content, this is not the full story. Recall that according to Wilson and Sperber (2004, p. 615), overall utterance interpretation involves three subtasks: identification of explicit content, intended contextual assumptions and intended contextual implications. So far, I have focused on the first of these subtasks, but in analysing interpretation of examples such as (54)-(59), we must consider the remaining two. These two tasks result in the derivation of implicatures,

⁹ Presumably the intonation and stress pattern associated with (60) will also vary with the discourse context. When there is more than one competing candidate referent, the demonstrative determiner will be emphasised ('THAT dog with three legs is called 'Lucky') in a way that is not necessary when there is only one three-legged canine present.

and therefore contribute to the implicit side of comprehension. As I have shown, the speaker may exploit procedural information to guide the hearer along the intended inferential path. If demonstratives encode procedural meaning, we would expect them to be used in this way to guide inferential processes contributing to both explicit content (i.e. reference resolution), and implicit content (i.e. implicatures). By applying relevance principles to the case of demonstratives, we can therefore incorporate into the analysis the role they may play in the inferential identification of implicatures. This is a role which has been largely overlooked in previous accounts, which have, for the most part, concentrated on the issue of how reference is resolved.

Consider again examples (54) and (55) ('Tartan was very popular in the 19th century. Prince Albert helped this/the trend by developing his own Balmoral tartan'). Although reference resolution is crucial to comprehension, a representation of the intended referent is already highly accessible in these cases, having been mentioned in the immediately preceding utterances. In both examples, the trend of wearing tartan is the most salient trend in the discourse context, and will therefore be tested first as a candidate referent by a hearer following the relevance theoretic comprehension procedure. This will happen regardless of whether or not encoded information about (discourse) proximity is provided. Since the identity of the intended referent is already taken care of, the speaker has available the option of using information about proximity encoded in the demonstrative determiner to create extra or different effects (so long as this does not interfere with correct assignment of reference). In cases of this type, the extra effort required to process the procedural information provided by the speaker is justified only if it yields an adequate range of additional inferential effects.

We then need to consider what form these additional effects might take. To avoid any suggestion that 'this' and 'that' are ambiguous, we want the procedural information encoded by each demonstrative to remain constant in all cases. Applying my earlier analysis, then, utterance (54) ('this trend') indicates that the trend of wearing tartan is closer to the deictic centre (in spatial, temporal or discourse terms) than some other competing trend. Although there is no competing trend in the discourse context, the hearer still presumes that the speaker is aiming at optimal relevance, and will infer the existence of a potential competitor or competitors from the use of the demonstrative form. Thus, while the nominal part of the referring expression is uniquely denoting in the discourse context, use of the demonstrative encourages the hearer to treat the intended referent as non-unique. This in turn evokes an implicit contrast effect which yields a wide range of weak implicatures. Such tacitly contrastive uses of

demonstratives might be analysed along similar lines to the stylistic effects resulting from repetition of words or phrases, as discussed by Sperber and Wilson (1986/95, pp. 219-221; Blakemore, 1992; 2008). Whilst some such uses contribute directly to the explicit content, others achieve relevance by increasing inferential effects, conveying suggestions of attitude or the speaker's degree of commitment to or involvement with the propositional content.

In this case, the contrast implicit in the use of a demonstrative stylistically highlights the wearing of tartan as one particular trend amongst many. Specific effects will depend on interaction between the implicit contrast and the wider discourse context, and may be affected by other factors such as intonation and tone of voice. For example, the speaker may be implicating that the fondness for tartan was just one of many trends (as opposed to serious preoccupations). Alternatively, it may just be a means of conveying a sense of continuity, indicating to the hearer that the subject matter has not changed and the focus remains on the same topic. Similar weak inferential effects are conveyed by examples (56)-(59). Here, both 'this' and 'that' are acceptable, because it is not the relative proximity of the referent itself that is relevant, but the implicit contrast with alternative possible referents. The choice of a demonstrative form over the definite article perhaps carries a tacit acknowledgement that the discussion focuses on one particular type of procedural information, where other types may be possible. The choice between 'this' and 'that' in such cases is likely to be driven by a variety of such subtle and interacting factors.

These subtle differences and apparent borderline cases support the hypothesis that the information about proximity encoded by the determiner can vary along several dimensions. They also reinforce the claim that proximity should not be characterised as an inherent property of an entity itself. Rather, it should be viewed as something that is computed from moment to moment, assigned to the referent online and subject to change as the discourse and discourse context develop. Proximity is a matter of degree which can vary as the physical, temporal and discourse contexts vary, leading to the wide range of contrast effects that we see in these examples¹⁰.

5.2.5 Revisiting the demonstrative categories

Recall that both Fillmore and Lakoff approached the issue of demonstratives by dividing the uses up into three broad categories: gestural/spatio-temporal,

¹⁰ It is only when these contrast effects clash with the overall sense of the utterance, as in the pareiasaur sentences discussed in chapter 4, that their different contributions become apparent.

anaphoric/discourse, and symbolic/emotional. I have suggested that these categories are both blurred at the edges and theoretically uninteresting, and argued that demonstratives should be analysed instead as encoding procedural information about relative proximity, broadly construed. I have tried to show how this procedural meaning may guide the hearer in resolving reference, and so contribute to the explicit content of an utterance, or it may contribute to the derivation of implicatures and other inferential effects. In this section, I reconsider the various uses identified by Fillmore and Lakoff, and show how they might fit into this relevance theory approach.

First, the examples classified by Fillmore and Lakoff as spatial / gestural involve use of the encoded procedural information to guide the hearer during reference resolution, and hence in the derivation of explicit content. The speaker indicates to the hearer that she expects him to pick out a certain referent from among two or more competitors, using the demonstrative form, perhaps accompanied by a physical demonstration, as a clue to her intended meaning. In these cases, substitution of a definite determiner for the demonstrative will be infelicitous.

With the discourse/anaphoric cases such as (12) and (13) ('Eat your greens: that's the solution to your problems' 'This is the solution to your problems: eat your greens'), application of the substitution test might seem to indicate that these function in a similar way. As with the gestural cases, substitution of the definite determiner for a demonstrative is unacceptable, as in (62). However, the cases are not really parallel, since (62) is not merely infelicitous but ungrammatical.

(62) *Eat your greens: the is the solution.

As has often been noted, the definite determiner differs from the demonstrative determiners in its inability to stand alone. Whilst 'this' and 'that' can function as either complex or bare demonstratives, 'the' must be accompanied by conceptual information or replaced by 'it', as in (63).

(63) Eat your greens: it's the solution.

Since in these cases, unlike the spatial / gestural ones, this substitution does not affect reference resolution, we can conclude that here the procedural information encoded by the demonstrative does not contribute to explicit content. Rather, use of the demonstrative in (12) introduces a contrast effect, and underlines the fact that the

solution is eating your greens, as opposed to doing anything else. Whilst the explicitly communicated meaning is the same in (63) ('Eat your greens: it's the solution'), the absence of a demonstrative leads to a loss of this extra, contrastive effect.

In discussing the data on demonstratives, Fillmore drew attention to the interaction between demonstrative forms and the tense of the utterance, as illustrated in examples (14)-(17) ('This/that has been an interesting lecture' 'This/that was an interesting course'). Notice that substituting the definite article for the demonstrative in these cases, as in (64) and (65), does not affect the explicit meaning:

(64) The course has been interesting.

(65) The course was interesting.

My analysis therefore predicts that the demonstrative will contribute to the implicitly communicated meaning in these cases: for instance by tacitly contrasting this particular course with other possible courses. Whilst in these examples substitution with the definite determiner is acceptable, the further choice between 'this' and 'that' is significant and directly affects the acceptability of the utterance. Again, this falls out naturally from the procedural analysis of demonstratives. A speaker who wishes to create an emphatic or contrast effect will choose a demonstrative form over a definite description. In many cases, it is the contrast effect that is relevant, and the choice between the demonstratives is marginal. In examples (56) and (57), for instance, ('This section focuses on what that/this procedural information might look like') both demonstratives are equally acceptable, and the choice between them might come down to subtle stylistic factors (e.g. the desire to avoid a repetition of 'this'). The speaker must choose one or other demonstrative determiner, but the choice itself may not be significant. However, in examples (14)-(17) the nature of the encoded procedure is significant since it interacts with the rest of the utterance. The use of the perfect continuous in (14) ('this has been...') maintains a connection with the present. As a result, the proximal form which indicates relative closeness to the deictic centre will be most appropriate, as it too implies a connection with the here and now. The use of the perfect aspect in (16) ('this was...') situates the event under discussion (i.e. the course) firmly in the past, and the distal form 'that' is more appropriate. These judgements support the hypothesis that demonstratives encode specific information relating to proximity, rather than just a vague instruction to draw a contrast. The contrast effect

arises from the interaction of the proximity information with the rest of the utterance and the discourse context.

A final challenge for any account of demonstratives is to shed some light on the wide range of disparate examples that fall under Lakoff's category of 'emotional' uses. Recall that Lakoff identified several types of 'emotional' use: for example, to create a sense of camaraderie or closeness, to replace the indefinite 'a', or to provide extra information about the subject of the previous utterance (see section 5.2.1.3, above). One advantage of a relevance theory approach is that it makes it possible to provide a unified account, since a wide range of different effects can arise from interaction between encoded meaning and discourse context, guided by the hearer's presumption of optimal relevance. I have argued that the relative nature of the encoded information about relative proximity can result in a contrast effect. I will now consider how this effect can be worked out in different ways in different contexts, to yield different types of weak implicatures, stylistic or poetic effects and attitudinal information. As Stephen Neale comments, 'uses of demonstratives upon which they are no more than stylistic variants of definite descriptions are legion' (2007, p.105).

Examples (22) and (25) illustrate one of the types of emotional use of 'this' and 'that' discussed by Lakoff (1974).

(22) A: I see there's going to be peace in the mideast.

B: This Henry Kissinger really is something!

(25) How's that throat?

In both cases, use of a demonstrative form is claimed to create a sense of camaraderie, intimacy or vividness. I want to suggest that this effect results from communication of the speaker's attitude to the referent, a possibility that has been overlooked in previous accounts. In neither case is the procedural meaning encoded by the demonstrative needed for reference resolution. To be optimally relevant, the utterance must therefore give rise to extra or different effects from those that would have been achieved in the absence of the determiner, as in the non-demonstrative versions in (66) and (67).

(66) Henry Kissinger really is something!

(67) How's your throat?

In all these cases, a representation of the intended referent is highly accessible to the hearer, and information about the proximity of the referent to the speaker is not required for reference resolution. In (22), the hearer will therefore look for extra or different effects to offset the effort of processing this otherwise superfluous linguistic material¹¹. Use of the proximal demonstrative determiner suggests an implicit contrast between the intended referent and other potential competitors (e.g. ‘that Henry Kissinger’, ‘the other Henry Kissinger’). Although in this discourse context, ‘Henry Kissinger’ is presumably uniquely denoting, the use of ‘this’ is naturally interpreted as implicitly contrasting different aspects, conceptions or even time-slices of Henry Kissinger. In this way, the speaker can communicate her attitude towards a particular aspect of Henry Kissinger or his behaviour. Although the attitude overtly indicated by use of the proximal demonstrative is one of emotional closeness, (22) may be interpreted as sincere or ironic depending on the discourse context, and the speaker’s intonation is likely to guide the hearer further on whether the speaker is associating herself with, or distancing herself from, the indicated attitude.

As Lakoff notes, use of the distal demonstrative in (25) (‘How’s that throat?’) creates a ‘sense of intimacy’ rather than distance between the doctor and her patient. How is this effect to be explained? My procedural analysis of demonstratives again suggests that it must be linked to some tacit contrast it evokes. In the first place, use of a demonstrative form sets the throat in question apart as of special interest to the doctor (in contrast with other throats), and indicates that both doctor and patient have discussed it before. In Cornish’s terms, “‘that’ functions interactively to create a sort of solidarity between discourse participants, establishing common ground’ (Cornish, 2001, p. 305). Again we return to Neale’s observation that use of that ϕ ‘draws attention to the fact that not any old ϕ will do’ (2007, p.103).

Uses of ‘this’ in place of the indefinite article ‘a’ make up Lakoff’s second category of ‘emotional’ uses, with examples such as (23) (‘There was this travelling salesman and he...’). Although cross-linguistic evidence (e.g. Ionin, 2006) suggests that indefinite ‘this’ does not necessarily belong in a unified account with definite demonstratives, my procedural approach does seem to have something to say in such

¹¹ This example is complicated slightly by the fact that ‘The Henry Kissinger’ would be infelicitous in this utterance (although ‘The Henry Kissinger we know...’ would be fine). Whilst both definite and demonstrative determiners share an encoded procedure that encourages the hearer towards a definite reference, this is not needed in and of itself in this case. However, a determiner of some form is required to carry the proximity information, and so create the extra effects.

cases. If the speaker had merely wanted to identify the intended referent, the most straightforward way would have been to use the indefinite article, as in (68).

(68) There was a travelling salesman, and he...

The hearer is therefore entitled to expect more or different effects from the use of the proximal demonstrative. According to Lakoff, this use 'gives greater vividness to the narrative' and 'involve[s] the addressee in it more fully' (1974, p. 347). On my account, this effect can be explained in the following way. Once again, the intended referent is presented as close to the deictic centre. As a result, despite the fact that the travelling salesman is being introduced into the discourse for the first time, the hearer is encouraged to see him as emotionally closer or more special to the speaker than other potential referents. What creates the sense of intimacy between speaker and hearer is the fact that he is expected, without any further help from the speaker, not only to realise that she has a certain range of competitors in mind, but to recognise which of them she is talking about. It is notable that there is no comparable use of 'that', and indefinite 'this' is mainly restricted to casual or informal registers, where implied intimacy between speaker and hearer would not be inappropriate.

Lakoff's final example of an 'emotional' use is illustrated by (24).

(24) 'Don't lie to me' said Dick. This was a man who had twice been convicted of perjury.

According to Lakoff, this emotional use 'provides additional information about [the referent] pertinent to the subject-matter of the prior sentence' (1974, p. 346). I want to suggest that this analysis overlooks an important aspect of such uses. In utterance (24), the content of the direct quotation ('Don't lie to me') is in sharp contrast with the description of the speaker that follows. The fact that Dick has said 'Don't lie to me' might lead the hearer to assume that Dick is an honest person who values truth. The next sentence provides information which is not merely 'additional', as Lakoff puts it, but contrasts strongly with this assumption. The procedural information encoded by 'this' precisely encourages the hearer to look for such a contrast, and achieves relevance thereby. In this way, the speaker not only guides the hearer towards the intended interpretation, but also communicates her own attitude towards Dick and his utterance. By emphasising the contrast, she communicates her intention that the utterances be seen

as carrying conflicting implications, and that this conflict is relevant enough to be worth the hearer's attention.

In this section, I have tried to show how the various uses of demonstrative forms can be analysed using the relevance theory notion of procedural meaning. Interpretation of utterances containing these items is guided by the relevance-theoretic comprehension procedure: the hearer follows a path of least effort, and stops when his expectations of relevance are satisfied. The procedural information encoded by the determiners narrows the search space and encourages him to look for inferences in certain directions. In some cases these inferences will contribute to reference resolution, and hence to explicit content, and in others to the derivation of implicatures.

My focus in this section has been on the English demonstrative system. Various other languages, including Greek, Japanese and Spanish, employ a three-way system which distinguishes proximal, distal and medial referents, while other systems may make still further distinctions (Diessel, 1999; Levinson, 2004). By treating proximity as a relation that is computed online as the discourse progresses, it should be possible to incorporate these further distinctions into a procedural approach.

An advantage of approaching the data using the relevance-theoretic notion of procedural meaning is that it becomes possible to provide a unified account of both bare demonstratives and complex demonstratives. The procedure remains the same in both cases, and the demonstrative contributes to explicit meaning in some cases, and to implicit meaning in others.

In sum, demonstrative determiners are a means by which the speaker may make her utterance optimally relevant. They may narrow the set of potential referents to a point where a representation of the intended referent is the most accessible one for the hearer, or they may be highlighting devices used to increase the accessibility of the intended referent for the same purpose. When a representation of the referent is already highly accessible to the hearer, the speaker may use a demonstrative form to achieve extra or different inferential effects. This may simply involve highlighting a perceived contrast which gives access to a range of implications or implicatures, or it may encourage the derivation of weak attitudinal implicatures. As I have tried to show, these effects are varied and wide-ranging. However, they can all be traced back to the encoded proximity information, its interaction with the context, and the speaker's goal of achieving optimal relevance.

Given my analysis of the 'emotional' uses of demonstratives, I turn again, in the next section, to the so-called 'special uses' of referring expressions identified by Ariel

(1990). Just as the seemingly disparate uses of demonstratives can be unified under a relevance-based account, so too, I will argue, can these special uses. Then, in the final section of this chapter, I will consider how stressed and unstressed pronouns might be treated on a relevance-based account of referring expressions. This will bring me closer to my ultimate goal of analysing null subjects in English within a general relevance-based account of referring expressions and considering how they contribute to both explicit content and implicatures.

5.2.6 'Special uses' revisited

In chapter 3, I discussed what Ariel (1990) calls the 'Special Uses' of referring expressions. These fall into three broad categories, illustrated in (69)-(71):

(69) That holiday we spent in Cyprus was really something, wasn't it?

(70) When John came home this afternoon, the son-of-a-bitch broke all the windows.

(71) This Henry Kissinger really is something!

According to Ariel, in (69) the intended referent (a holiday in Cyprus) is treated as accessible when it is not; in (70) a highly accessible referent ('John') is referred to using a low accessibility marker ('son-of-a-bitch'); and in (71), a referent ('Henry Kissinger') is picked out using a marker of higher accessibility ('this') than might be expected. Although Ariel claims that such inappropriate uses encourage the hearer to derive extra contextual implications, I pointed out some problems with her analysis. The most crucial problem is the question of how the hearer decides whether to take the level of Accessibility signalled by the referring expression at face value, or whether to treat it as unusual, and therefore go in search of extra implications. There is also a question about how the hearer decides whether he is expected to derive positive or negative effects from the inappropriate use of a particular term.

Ariel's illustrative examples in (69)-(71) are parallel to examples already discussed in my relevance-based account above, and I revisit them here to briefly illustrate how they would be dealt with on my account.

In all three examples, (69)-(71), the intended referent is the one that is most highly accessible to the hearer and will therefore be tested first. Compare the versions above with the alternatives in (72)-(74).

(72) The holiday we spent in Cyprus was really something, wasn't it?

(73) When John came home this afternoon, he broke all the windows.

(74) Henry Kissinger really is something!

It seems that, in order to resolve reference and derive the explicit content of the utterance, the extra information provided by the demonstrative determiner or anaphoric epithet in (69)-(71) is not required. In each case, the alternative versions in (72)-(74) are acceptable and lead to the same assignment of reference and the same explicit content. Hence, the choice of referential expression in (69)-(71) must contribute to the implicitly communicated meaning, and my relevance theoretic account helps to explain how it does so. In section 5.2.6.2, I showed how the procedural information encoded by 'this' and 'that' can contribute to what is implicitly communicated by an utterance, and in chapter 4 I discussed a parallel case to (70) (John_i went into town for his lunch. The rascal_i was late back to the office). Thus, such uses need no longer be considered special or exceptional, but fall out naturally from my account. Furthermore, the fact that, according to relevance theory, the overall interpretation of an utterance crucially depends on the discourse context in which it is uttered, allows us to explain why an utterance of (71) might communicate either admiration or derision, depending on the contextual assumptions accessible to the hearer (perhaps guided by non-linguistic cues such as facial expression or tone of voice).

To conclude, an advantage of my relevance-based account is that it treats the contribution that an expression may make to implicatures (as well as to explicatures) as fundamental. On this approach, the effects described by Ariel are no longer special, but simply fall out naturally from the relevance-theoretic comprehension procedure, as part of the overall interpretation of the utterance.

In this section I have shown how the distributional differences between definite descriptions and complex demonstratives, on the one hand, and distal and proximal demonstratives, on the other, can be explained within a relevance theoretic analysis of referring expression. I have also argued that Ariel's 'special' uses are not special after all.

The final distinction that Gundel and colleagues regard as problematic for relevance theory is between stressed and unstressed pronouns. In the next section, I apply my analysis to these data, and argue that the use and interpretation of such

expressions falls out naturally from my account, which in turn makes the appeal to notions of encoded Accessibility or Givenness superfluous.

5.3 Stressed and unstressed pronouns

5.3.1 Relevance and accessibility

I have argued that we should abandon the idea that information about ‘Accessibility’ or ‘Givenness’ is an inherent property of referring expression types, which has to be linguistically marked or encoded. Rather, the accessibility of various representations of a potential referent depends on a range of factors that vary from context to context. The most accessible representation of a referent is simply the first one that a hearer with a relevance-oriented cognitive system comes across when following the path of least effort. According to the relevance-theoretic comprehension procedure, hearers will test interpretive hypotheses in order of accessibility. A speaker aiming at optimal relevance will therefore use the form of referring expression that should make a representation of the intended referent the easiest one to retrieve. The notion of ‘accessibility’ mentioned in the relevance theoretic comprehension procedure does not, therefore, need to be defined in terms of distance, competition or cognitive status, along the lines proposed by Ariel and Gundel. Although each of these factors may play a role in the interpretation of certain utterances, on certain occasions, they are not essential to the working of the comprehension procedure, and do not play a crucial theoretical role. Factors which may influence the accessibility of representations of potential referents may be found in the physical, linguistic or discourse context, or any combination of these. In section 5.3.2, I briefly consider some factors that have been shown to affect accessibility, before returning to the analysis of stressed versus unstressed pronouns in section 5.3.3.

5.3.2 Factors affecting accessibility of interpretations

Matsui (2000) discusses several factors that may contribute to the accessibility of a representation of a potential referent, and cites experimental data from existing studies of these factors. In her view, all these factors should be ‘integrated into one higher level factor, namely, processing effort’ (Matsui 2000, p. 52). This conclusion brings us back, to the relevance theoretic comprehension procedure and my argument that use of referring expressions is ultimately driven by considerations of relevance. In this section, I will outline some of the factors discussed by Matsui. To reiterate, this is not

an attempt to define what it means to be accessible, but rather to outline some of the factors which have been shown to affect the accessibility of certain interpretations in certain contexts.

5.3.2.1 Order of mention

Drawing on ‘centering theory’ (Joshi & Weinstein, 1981; Grosz, Joshi, & Weinstein, 1995; Walker, Joshi, & Prince, 1998), and integrating experimental results from Gernsbacher & Hargreaves (1988), Matsui discusses the effect of word order on accessibility. In experimental studies, words appearing earlier in an utterance were responded to more quickly than those appearing later in the same utterance. This result persisted even when the grammatical relations and semantic roles were varied. The results of these experiments revealed that first-mention participants were responded to faster than second-mention participants, and furthermore that first-mention non-subjects were responded to faster than second-mention subjects. Thus, order of mention seems to influence the accessibility of a potential referent, and the ‘higher accessibility of the first-mentioned item is not attributable to their subjecthood’ (Matsui 2000, p. 57).

5.3.2.2 Syntactic position

Although order of mention seems to affect accessibility regardless of grammatical role, the syntactic structure of an utterance does appear to influence the order in which potential referents will be tested. Experimental work by Matthews and Chodorow (1988) suggests that ‘there is a left-to-right top-down breadth-first search of the prior clause for an appropriate antecedent’ (p.256). McKoon et al (1993) consider the influence of syntactic position from a slightly different perspective, and argue that ‘the more affected a discourse entity is by the action of the verb...the more prominent or salient will be its position in a discourse’ (p.595). So, for example, a participant in direct object position would be more accessible than one in indirect object position, which in turn would be more accessible than a participant presented in a prepositional phrase.

5.3.2.3 Semantics of the main verb

Caramazza et al (1977), following Garvey & Caramazza (1974) and Garvey et al. (1976), introduce the notion of ‘implicit causality’ as a possible factor affecting the accessibility of potential referents in a discourse context, and they claim that in certain constructions, the semantics of some verbs can bias the interpretation of pronouns

towards either the first or second noun phrase in the sentence as antecedent. Consider the utterances in (75) and (76).

(75) John telephoned Bill because he wanted some information.

(76) John telephoned Bill because he withheld some information.

They predict that verbs such as ‘telephone’ bias interpretation of the pronoun towards the first noun phrase in the sentence, so that sentence (75), which is plausible on this interpretation, will be processed faster than (76), which is not. Their experiments supported this claim. It was found that verbs such as ‘call’, ‘sell’ and ‘join’ pattern with ‘telephone’, whereas ‘fear’, ‘blame’ and ‘kill’ bias interpretation towards the second noun phrase in the sentence.

I would like to suggest that this sort of bias can be explained in terms of relevance theoretic notions, rather than treated as an arbitrary fact about the semantics of the verb. The reason for fearing, blaming, killing someone is more likely to lie with the person being feared, blamed, killed, whilst calling, selling, joining and telephoning are more readily seen as agent driven actions. Ultimately, the hearer will settle on an interpretation which makes the utterance optimally relevant: it is easier to construct a context in which Bill’s withholding of information leads John to phone him than one where John’s withholding of information prompts the call. Withholding suggests intent, and it is hard (though not impossible) to construct a context where John’s deliberately withholding information explains his telephoning Bill.

5.3.2.4 Parallel function

Sheldon (1974), again considering the resolution of pronouns, introduces the idea of parallel function. She hypothesizes that ‘In a complex sentence, if coreferential NPs have the same grammatical function in their respective clauses, then that sentence will be easier to process than one in which the coreferential NPs have different grammatical functions’. Thus, in a sentence such as (77), ‘he’ will usually be interpreted as referring to John rather than Bill.

(77) John hit Bill and he kicked Sarah.

However, this notion of parallel function seems to interact with stress, as in (78), where ‘he’ is now naturally understood as referring to Bill.

(78) John hit Bill and then HE kicked Sarah.

There is also evidence that it interacts with the semantics of both the verb and the conjunction. For instance, Ehrlich (1980) shows that the ‘implicit causality’ described above is affected when ‘because’ is replaced by ‘but’. Consider (79) and (80).

(79) John feared Bill because he talked loudly.

(80) John feared Bill but he talked loudly.

Whilst in (79) the most accessible interpretation for the pronoun seems to be Bill, in (80), ‘he’ is more naturally interpreted as referring to John.

5.3.2.5 Accessibility of contextual assumptions

A problem with trying to account for reference resolution purely by considering factors that influence the relative accessibility of potential referents is that these various features interact both with one another and with the discourse context. Matsui (2000) takes the discussion forward by returning to the relevance theory notion of processing effort. Processing effort is affected not only by the accessibility of candidate referents, but by the accessibility of contextual assumptions. She comments that ‘when two candidate referents are roughly equally accessible, it is the accessibility of contextual assumptions that plays the decisive role’ (p. 88). I want to take her argument further, and suggest that accessibility of contextual assumptions can affect the overall interpretation even if the hearer has not reached a stalemate as far as the accessibility of the potential referents is concerned. Consider the following examples.

(81) John visited Bill because he wasn’t feeling well

(82) John visited Dr Smith because he wasn’t feeling well

(83) Dr Jones visited Bill because he wasn’t feeling well

These three utterances differ only in the specifics of the particular proper names used ('John', 'Bill', 'Dr Smith', 'Dr Jones'). However, even if it is clear that 'John' co-refers with 'Dr Jones' and 'Bill' co-refers with 'Dr Smith', these proper names provide easy access to contextual assumptions which are likely to lead to different assignments of reference to the pronoun 'he'.

In all three utterances the pronoun 'he' could in principle refer to either Dr John Jones or Dr Bill Smith. In (81), both interpretations are compatible with highly accessible contextual assumptions: for instance, it is plausible that John visited his friend Bill because John wasn't feeling well and wanted advice or to be cheered up, and it is plausible that John visited his friend Bill because Bill wasn't feeling well¹². However, the choice of proper names in (82) and (83) interacts with highly accessible contextual assumptions about doctors and the reasons for visiting them, to make particular interpretations strongly preferred in these cases. In each case, the highly accessible assumption that people who don't feel well go to see doctors leads to a preferred interpretation where the participant who is feeling ill is the one who is not the doctor. In (83), the influence of factors such as order of mention and syntactic position is overridden by the overall search for a relevant interpretation.

As Wilson (1992, p. 186) puts it, 'reference assignment does not depend on accessibility of referents alone, but is also influenced by the need to obtain an interpretation that satisfies some criterion of pragmatic acceptability'. In each case, the use of the connective 'because' encourages the hearer to interpret the second conjunct as offering an explanation for the first. Thus, reference will be assigned to 'he' in such a way as to combine with easily accessible contextual assumptions to offer such an explanation.

5.3.2.6 Referring expressions and accessibility of interpretations

In this subsection I have discussed some of the factors affecting accessibility outlined by Matsui. These factors have been shown to have a significant influence on the relative accessibility of candidate referents. In combination, these factors affect processing effort, and hence the accessibility of interpretations: either by directly affecting the accessibility of a mental representation of the intended referent, or indirectly by affecting the accessibility of different contextual assumptions, on the one hand, and the hearer's expectations of relevance, on the other. However, I do not see

¹² Perhaps the latter interpretation is more likely to be preferred, as the assumption that sick people sometimes receive visitors may be the most highly accessible.

the above list of factors as by any means exhaustive. Rather, the factors identified by Matsui indicate some of the many ways in which the linguistic form of an utterance can affect the processing effort demanded of the hearer. Processing effort is, of course, not the full story when it comes to defining relevance, and my goal is not to define accessibility in terms of a list of factors of the type discussed above. Rather, I start from the assumption that at any given point in a discourse, referential candidates will be more or less accessible relative to each other, and to the accessibility of contextual assumptions in that particular discourse context. My account of referring expressions takes seriously the claim of relevance theory that interpretations are tested in order of accessibility. The speaker must therefore undertake some ‘mindreading’ in order to assess the accessibility to the hearer of different candidate referents, and construct her utterance accordingly, while monitoring the hearer’s reactions as the utterance proceeds and making repairs if necessary. Of course, mindreading, like utterance interpretation itself, is fallible, and it is not surprising that misunderstandings and misinterpretations can and do occur. As Sperber and Wilson put it, what is surprising when people engage in ostensive-inferential communication is not failure, but success.

In sum, a speaker aiming at optimal relevance will construct her utterance so that the first satisfactory interpretation to occur to the hearer (i.e. the first one that satisfies the expectations of relevance raised by the utterance) will be the intended one. Choosing an appropriate referring expression is just one aspect of this process.

5.3.3 Stressed and unstressed pronouns: a relevance-theoretic account

Bearing in mind the factors affecting the accessibility of candidate referents outlined by Matsui, I return to the examples of stressed and unstressed pronouns from chapter 3.

(84) Jane kissed Mary and then she kissed Harry.

(85) Jane kissed Mary and then SHE kissed Harry.

When the pronoun ‘she’ is unstressed in example (84), it is most naturally interpreted as referring to Jane. However, by stressing the pronoun, as in (85), the speaker can change the interpretation so that the preferred interpretation is one on which Mary is kissing Harry.

Considering (84) first, the procedural information encoded by the pronoun ‘she’ narrows the set of potential referents to just those that are female¹³, i.e. the set of all female entities. Within that set, Jane and Mary are likely to be highly salient by virtue of having just been mentioned. However, in (84), the pronoun ‘she’ is most naturally interpreted as referring to Jane, and so, on the assumption that the hearer is following the relevance theoretic comprehension procedure, this interpretation must be the more accessible of the two. Matsui’s overview of contributing factors might give us some clue as to why a representation of Jane might be more accessible at that point in the discourse. Jane is the subject of the sentence and the first-mentioned candidate referent, and both these factors have been shown to contribute to the accessibility of a candidate referent. The accessibility of a representation of Jane may also be boosted by parallel function, since both ‘Jane’ and ‘she’ play the grammatical role of subject in their respective clauses. For these reasons, ‘Jane’ is likely to be tested first, and if the resulting overall interpretation seems likely to satisfy the hearer’s expectations, he need look no further.

In (85), the same procedure applies. The hearer follows the path of least effort, and stops when he finds an interpretation that satisfies his expectations of relevance. However, a stressed pronoun puts the hearer to more effort than an unstressed one. According to Wilson and Wharton (2006) ‘if two stress patterns differ in the amounts of processing effort required, the costlier pattern should be used more sparingly, and only in order to create extra or different effects’ (2006, p. 1567). If the intended interpretation were ‘Jane’, then the speaker could have uttered (84). The very fact of stressing the pronoun will therefore prompt the hearer to search for an alternative interpretation to justify the extra effort required. He therefore proceeds along the path of least effort and tests the next, still highly accessible, potential referent – in this case Mary. The different interpretation justifies the extra effort by yielding a different explicit content.

However, as I have shown above, reference resolution and the identification of the intended explicit content is only one of the three subtasks in the hearer’s search for the intended overall interpretation. In an example such as (84), both possible interpretations (with ‘she’ as co referential with either Jane or Mary) are equally plausible in the (limited) discourse context. When all other things are equal, the type of factor listed by

¹³ Since what is relevant about these examples for present purposes is the difference made by the absence or presence of stress, I abstract away from the details of the procedure here. In chapter 6 I will examine the nature of the procedures encoded by personal pronouns in more detail.

Matsui as affecting the accessibility of potential references may be seen as playing a decisive role. However, consider the more complicated examples in (86) and (87).

(86) John criticized Bill because he talks too much.

(87) John criticized Bill because HE talks too much.

As always, the hearer of these utterances will test interpretive hypotheses in order of accessibility. However, although in both utterances the syntactic status of 'John' (subject position, first mention etc) may seem to make him the most accessible potential referent for the pronoun 'he', this interpretation of (86) will fail to satisfy the hearer's expectation of relevance. As noted above, the choice of the connective 'because' encourages the hearer to expect the following clause to provide some explanation for John's having criticized Bill. In fact, two different possible explanations come to mind, so that (86) may have the interpretation in either (88) or (89).

(88) John criticized Bill because John talks too much.

(89) John criticized Bill because Bill talks too much.

I suggest that a context in which Bill's verbosity results in John's criticism is likely to be much more accessible to a typical hearer than one in which John's criticism is motivated by his own excessive talking. Whereas in example (84), both potential interpretations are equally plausible, and so the first one tested (in this case Mary) is the only one that satisfies the hearer's expectations of relevance, in (86) all other things are not equal. These examples illustrate how the three sub-tasks in overall interpretation take place in parallel and can interact with the speaker's choice of referring expression. To take the argument one stage further, consider (90).

(90) John criticized Bill because he was in a bad mood.

Without the use of intonational cues, this comes much closer to being truly equivocal between the two readings. We can easily imagine a context in which Bill's being in a bad mood led to John's criticism, and we can also imagine that John being in a bad

mood could lead him to criticize anyone who crosses his path – in this case Bill. Here, further contextual assumptions are needed to resolve reference one way or the other.

As these examples show, stress can contribute to the explicit content of an utterance by affecting reference resolution. However, just as with the other types of referring expression discussed above, stress can also contribute to the implicit phase of interpretation. Suppose (91) is uttered at a party as Mary and Jane both notice that Sue has walked in.

(91) Mary: Nobody told me SHE was coming.

Here, the procedural information encoded by the pronoun ‘she’ narrows the set of potential referents to just those that are female. However, since the fact that Sue has just walked in is mutually manifest to both speaker and hearer, Sue is likely to be more salient to them than any other female present, and would probably be the first referent to be tested even if the pronoun had been unstressed. Thus, adding stress to the pronoun once again encourages the hearer to look for extra or different effects to justify the extra processing effort. In this case, however, all the other females in the discourse context are equally salient, and there is no other potential candidate whose representation is accessible enough to be worth testing. If one of these were indeed the intended referent, the speaker would have to narrow the set in such a way as to make this female more salient than the others. The hearer of (91) will therefore have to justify the extra effort not by altering the reference of ‘she’, but by deriving extra effects during the inferential phase of interpretation. Thus, by stressing ‘she’, Mary invites the hearer to assume that she intends to do more than just secure reference: for instance she may want to communicate her own attitude towards Sue. How exactly the interpretation will go is likely to depend on the exact intonation of the utterance, the speaker’s body language, facial expression and so on.

In this section, I have tried to show that my relevance-based account can distinguish between stressed and unstressed pronouns. This provides an answer to Ariel’s remaining objection to a relevance-only account of the use of referring expressions, and allows us to dispense with the notion that referring expressions encode information relating to Accessibility or cognitive status.

5.4 Conclusion

Ariel (1990, p. 82) considers the possibility that her Accessibility scale could be ‘replaced by a relevance theory account, possibly rendering Accessibility theory redundant’, but concludes that considerations of relevance alone cannot distinguish between each type of referring expression. In this chapter I have tried to show that Blakemore’s notion of procedural meaning provides us with a means to do just this.

In discussing the ‘special uses’ of referring expressions, Ariel acknowledges that there is also the ‘fascinating subject of what guides speakers in choosing a specific expression given a variety of expressions identical in terms of degree of Accessibility’ (1990, p. 198). She concludes that ‘most of these problems should be handled by Relevance directly, rather than by Accessibility’ (1990, p. 198). In this chapter, I have built on the ideas introduced in chapter 4 to show that we can take this approach one step further and use relevance theory (including the notion of procedural meaning) to motivate the distinctions between each of the different forms of referring expression, and to account for the so-called special uses without the need for externally motivated scales or hierarchies. In doing so, I have also argued that the different forms of referring expression affect the overall interpretation of an utterance in a variety of ways, and that this reveals Ariel’s and Gundel et al.’s treatments of reference resolution to be inadequate.

As my ultimate goal in this thesis is to provide a cognitively-orientated pragmatic account of null subject utterances in English, I aim to show that they are a legitimate means of securing reference in certain discourse contexts, and that their use is unremarkable when certain conditions obtain. In the next chapter, I argue that, as with other referring expressions, null subjects may contribute not only to reference resolution, but to the overall interpretation of an utterance. Having established a general account of referring expressions, I am now in the position to show how null subjects might fit into my approach. In chapter 6, I return to the null subject data to do just this.

Chapter 6: Null Subjects Revisited

6.1 Introduction

In the first part of this thesis, I discussed the phenomenon of null subjects in English diary-style texts, and in later chapters I presented a relevance-based analysis of referring expressions. In this chapter I combine these two themes to offer a relevance-based analysis of null subjects. I begin by considering how zero subjects in English might fit into the existing accounts of referring expressions discussed in earlier chapters. In section 6.3, I consider how my relevance-based approach to referring expressions might be extended to deal with overt pronominal forms. Finally, in section 6.4, I consider the implications of the resulting account for the treatment of null subjects themselves. I try to show that null subjects, like other referring expressions, contribute to the overall relevance of an utterance, and that they may do this either by minimising the hearer's effort, or by creating extra or different effects whilst allowing for any constraints imposed by the speaker's preferences and abilities. As such, they fit naturally into a relevance-based account of utterance interpretation in general, and referring expression use in particular.

6.2 Null subjects as referring expressions

6.2.1 Accessibility and null subjects

According to Accessibility theory (Ariel, 1990) both pronouns and gaps are markers of high accessibility. Ariel devotes a chapter to looking in more detail at zero subjects. She begins by noting that not all zero subjects are alike, and that we should not assume that they 'constitute a unitary phenomenon' (1990, p. 107). On the 'commonsensical' assumption that the subject of a sentence is 'often interpreted via the AGR element', she suggests that 'we should examine the nature of the AGR element' in order to 'determine the degree of Accessibility associated with each AGR type' (1990, p. 107).

From this starting point, she goes on to provide an in-depth study of zero subjects in Hebrew, and uses this to construct a Hebrew-specific hierarchy which incorporates full pronouns, cliticized pronouns and AGR elements. Turning her attention to AGR in English, she claims that it does not differ significantly from the Hebrew present tense, which 'in general rarely allow[s] zero subjects' (1990, p. 122). Noting that zeros may

occur when the intended referent (often the speaker or addressee) is particularly salient, she concludes that zeros 'mark extremely high Accessibility' (p. 130). This high level of Accessibility may be a result of either syntactic or discourse factors:

The condition imposed on all extremely High Accessibility Markers is that the mental representations they evoke be highly accessible to the addressee. The source for this status of high Accessibility is irrelevant, although languages do differ with regard to what this source may be in the unmarked case (Ariel 1990, p. 130).

Although this approach enables Ariel to incorporate English null subjects into the Accessibility theory account of referring expressions, it does little to explain why these null subjects should occur in the first place. In her analysis of Hebrew, Ariel finds that third person pronouns and zero subjects 'occur in perfectly complementary distributions' (p.121). However, this is not the case in English. In diary-style examples such as (1), overt third person pronouns can be substituted for zero subjects in all cases without affecting the acceptability of the utterance, as in (2):

- (1) Probably wants you to reply to his message.
- (2) He probably wants you to reply to his message.

This distributional fact has implications for the treatment of null subjects in English. Ariel's analysis of the Hebrew AGR is partly motivated by the desire to discover what governs speakers' preferences among the various choices of pronominal forms (full pronouns, clitics and zeros). When the variants occur in complementary distribution, as in Hebrew, it is reasonable to appeal to a distinguishing factor such as the notion of encoded Accessibility to explain the distribution. However, the English null subject data pattern differently. Whilst Ariel's claim that the variations in Hebrew speakers' choice of referring expressions 'are not random' carries over to the English data, her further claim that the variations can be 'motivated by Accessibility theory quite straightforwardly' (1990, p.116) is not so easily maintained.

I do think, however, that Ariel's discussion of 'special uses' of referring expressions may shed some light on the English null subject phenomenon. Her suggestion is that the higher the Accessibility marker used, the more the speaker indicates that she empathizes with the intended referent, and it seems that this point

could apply to English null subjects too. This aspect of Ariel's analysis is compatible with the fact that, in English, overt forms can always substitute for the nulls without affecting overall acceptability. It could be argued that by using an extremely high Accessibility marker where a high Accessibility marker would have sufficed, the speaker is encouraging the hearer to derive extra contextual implications, having to do, for instance, with her own attitude of empathy. In chapter 3, I discussed examples where Ariel claims that choice of referring expression contributes to an empathy effect. This acknowledgement that choice of referring expression might affect what is implicitly communicated by an utterance is crucial if we are to understand the (optional) alternations between overt and null referring expressions in the diary-style texts in English. However, the problems raised by an Accessibility theory analysis of these special uses (as discussed in 3.3.5) remain. In section 6.4, I offer a relevance-based analysis which eliminates the need for Accessibility theory, and therefore also removes these associated problems.

6.2.2 Accessibility and avoiding pronouns

Ariel (1990) makes some more general observations about the treatment of null subjects in Accessibility theory. In particular, she argues that Chomsky's Avoid Pronoun Principle (Chomsky, 1981; also see chapter 1, section 1.6.1) – to the extent that it applies – 'naturally falls out from the more general theory of Accessibility' (Ariel, 1990, p. 101). The principle, repeated here in (3), imposes 'a choice of PRO over an overt pronoun when possible' (Chomsky, 1981, p. 65).

- (3) Avoid Pronoun.

Compare the interpretation of (4) with that of (5), in which the pronoun could equally refer to John or to some other contextually salient male.

- (4) John would much prefer [his going to the movie].

- (5) John would much prefer [his book].

Replacing 'his' with PRO is an option in (4), but not in (5). The Avoid Pronoun principle explains why the possessive pronoun 'his' in (4) must be disjoint in reference

with 'John' (unless stressed). If the speaker had intended to refer to John, she should have used PRO, thus avoiding use of an overt pronoun.

According to Chomsky, the Avoid Pronoun principle 'might be regarded as a subcase of a conversational principle of not saying more than is required' (1981, p. 65). Ariel argues that treating it as a grammatical principle is problematic. She explains:

Given that speakers...opt for Optimal Relevance...if they can achieve the same goal with less effort (obviously zero is less of an effort than the pronunciation of an overt pronoun), why should they ever produce the overt forms?¹

Since overt pronouns do occur in *pro*-drop languages, Ariel concludes that the principle in (3) 'cannot be made responsible for the whole phenomenon of *pro*/pronoun distribution' (1990, p. 103). In her view, the Avoid Pronoun principle is both 'unnecessary and incorrect' (p. 104), and Accessibility theory can better account for the distribution of overt pronouns and gaps. As discussed above, in Ariel's framework, the speaker is sensitive to the Accessibility of the intended referent in the discourse context, and chooses a referring expression which signals the appropriate level of Accessibility. Whilst I agree with Ariel's suggestion that the Avoid Pronoun principle is 'more apparent than real', once again I want to question why Accessibility theory is needed, in addition to the general principles of relevance.

According to relevance theory, a speaker aiming at optimal relevance will formulate her utterance so that the first interpretation the hearer tests and finds relevant in the expected way will be the intended one. As we have seen, hearers test interpretations in order of accessibility. At the point in (4) where 'his' is interpreted, John is the most accessible potential referent, but if John had been the intended referent, a speaker aiming at optimal relevance should have uttered (6) instead:

(6) John would much prefer [PRO going to the movie].

¹ This claim reveals a possibly important problem with Ariel's understanding of relevance theory. The crucial type of effort involved in achieving optimal relevance is hearer's processing effort rather than speaker's production effort. As noted by Carston (2005), there are many possible factors which may 'affect the degree of effort a speaker is willing to expend in encoding her thoughts and articulating linguistic forms'. Although minimizing her own effort may be one such factor, it is likely to be integrated with a range of other goals, and is incorporated into the definition of optimal relevance under the heading of 'the speaker's abilities and preferences'.

The choice of the overt pronoun ‘his’ in (4) causes the hearer extra linguistic processing effort, and he is therefore entitled to look for more or different cognitive effects to justify this extra effort.² The easiest way to obtain such effects is to assign reference to another salient male in the discourse context.³ In this way, the disjoint reference predicted by the Avoid Pronoun principle falls out naturally from relevance theory.

Interpretations of the type just described are to be expected if (as I will go on to argue in more detail) pronouns encode procedural meaning, and procedural meaning saves the hearer effort by narrowing the search space for inferential processes (including reference assignment). Overt realisation of the pronoun would only be worthwhile if the procedural meaning saves the hearer from going down the wrong inferential path. In example (6), which lacks an overt pronoun, John is already the most accessible referent at the point where PRO is interpreted, and will therefore be tested first. Since (6) is fully grammatical, there is no reason to expect the hearer to go to the extra effort of processing an overt pronoun. Indeed, if the overt pronoun form were used, the hearer would be likely to assume that the extra effort demanded was not gratuitous, and therefore look for alternatives to the co-referential interpretation. Where the null pronoun (PRO) is not a grammatical option, as with ‘his book’ in (5), the overt pronoun should have two possible readings, one on which it refers to John, and the other on which it refers to some other salient male in the discourse context. The hearer will then settle on the interpretation which most easily combines with accessible contextual assumptions to yield an overall interpretation that is relevant in the expected way.

6.2.3 The Givenness hierarchy and null subjects

The other major pragmatically-oriented approach to referring expressions, based on the Givenness hierarchy, also offers to shed some light on null forms in English. Gundel, Hedberg and Zacharski (1993) examine the distribution of referring forms in five languages (Chinese, English, Japanese, Russian and Spanish), and conclude that ‘the six statuses on the Givenness Hierarchy are adequate for describing appropriate use’ in those languages (1993, p. 284). Although the status hierarchy itself is intended to be adequate cross-linguistically, the correlation between referring forms and cognitive statuses is seen as varying between languages. Notably, English is the only language in the survey which does not utilize zero subjects as standard. For each of the other

² This follows a similar pattern to my analysis of stressed versus unstressed pronouns in chapter 5.

³ If the pronoun is stressed, (John would much prefer HIS going to movie), John again becomes a possible referent. The extra effort to which the hearer is put is justified by a contrastive interpretation.

languages, the zero form is seen as requiring that the referent be 'in focus'. How, then, might a Givenness hierarchy analysis of referring expressions deal with the diary-style zero subjects in English and other non-null subject languages?

It seems that there are two possibilities. First, we might treat the zero subjects as signalling some sort of 'super-in-focus' status to the hearer. This would enable the speaker to distinguish between several competing 'in focus' potential referents in cases where one is more salient than the others, and so could therefore be classed as 'super-in-focus'. Alternatively, it could be that when zeros alternate with overt pronouns, the contrast has nothing to do with the cognitive status of the referent, and instead serves some other purpose in the discourse.

Both options are problematic for a Givenness-based account of referring expressions. According to the first possibility, zero subjects in English would be analysed as representing some higher cognitive status not otherwise allowed for in the original scale. This raises important questions about the universality of the approach. It would mean that the six statuses identified by Gundel et al (1993) are not, after all, cross-linguistically adequate. If we take the step of adding this language specific extra distinction, might there then be room for further distinctions, and if so, at what point does the hierarchy cease to involve language specific manifestations of universal principles, and instead just become language specific? Furthermore, if we introduce the idea of a 'super-in-focus' cognitive status, we might then ask why this does not seem to be needed (or even possible) in languages such as Japanese and Russian, where zeros already signal 'in focus'? Finally, a 'super-in-focus' status could only be justified in English if it is needed to allow the speaker to distinguish between competing 'in focus' candidate referents. However, in many null subject utterances, there is only one 'in focus' potential referent in the discourse context. This is particularly evident in examples taken from diaries, such as (4)-(9), where the only 'in focus' entity is the author of the diary. It is unclear how the use of a zero subject could aid reference resolution in these cases.

(7) Walked to Westminster hall (Pepys, 1660, p. 610).

(8) Sat there for $\frac{3}{4}$ hour (Woolf, 1902, p. 334).

(9) Went to the chemist to discreetly buy pregnancy test (Fielding, 1749, p. 117).

These examples bring us to a second possibility: that the alternation between overt and zero forms relates to something other than the cognitive status of the intended referent. The status ‘in focus’ in English would then be seen as correlating with both zero and overt pronouns, and the speaker’s choice of one over the other would be driven by some other factor, and would serve some other purpose in the discourse. This is exactly the position that my relevance-based account will take, and although it is not incompatible with the Givenness hierarchy per se, an approach along these lines would bring into question the motivation for the hierarchy in the first place. As discussed in the previous chapter, relevance theory seems able to account for the data that motivate each of the distinctions made by the Givenness hierarchy. Once zeros in English are added to the equation, it seems that any analysis of referring expressions must acknowledge the importance of factors influencing aspects of interpretation that go beyond mere reference resolution. Again, this is something that I have argued is missing from the existing accounts. It seems that the Givenness approach must appeal to a broader pragmatic framework such as relevance theory to do this. If relevance theory (or an alternative pragmatic framework) is needed to fully account for the use and interpretation of referring expressions, whilst at the same time being able to account for the distinctions made by the hierarchy, then it is unclear what advantage is gained from retaining the hierarchy.

6.3 Relevance theory and overt pronouns

6.3.1 Overt pronouns as referring expressions

The null subjects found in diaries and other abbreviated registers seem to have some referential content. As discussed in chapter 1, they alternate with overt pronouns throughout the texts in which they occur. In the previous chapter I argued that referring expressions contribute to the overall relevance of an utterance, and that they do this by encoding procedural and/or conceptual information which may contribute to the explicitly or implicitly communicated meaning. To understand how null subjects might be incorporated into this analysis, it is therefore necessary to consider in more detail how overt pronouns contribute to the overall interpretation of an utterance. It is this contribution that is missing in corresponding null subject utterances, and by understanding what it is, we can address the issue of why it seems to be optional in certain contexts.

6.3.2 Types of pronouns

Not all pronouns are alike. In a classic work, Evans (1980) distinguishes four categories of pronouns: deictic, anaphoric, bound and e-type. Deictic pronouns ‘make reference to an object...in the shared perceptual environment or rendered salient in some other way’, whilst anaphoric pronouns are interpreted as ‘coreferential with a referring expression elsewhere in the sentence’ (p.337). In the remaining two categories the pronouns take quantifier expressions as antecedents, bound, as in (10), and non-bound, as in (11).

(10) Every boy respects his teacher.

(11) Every man who owns a donkey feeds it.

In considering the phenomenon of null subjects, I restrict my analysis to the first two categories, which I will group together as ‘referring pronouns’.

I will also follow Heim & Kratzer (1998) in assuming that ‘there is no difference between anaphora and deixis’ (1998, p, 241):

Anaphora may often be viewed as reference to a contextually salient individual...it seems to differ from deixis only insofar as the *cause* of the referent’s salience is concerned (1998, p. 240).

This echoes Ariel’s (1990) proposal to do away with the referential-anaphoric distinction (see 3.3.1), and chooses to treat all referring expressions as anaphoric. On similar lines, Powell argues that ‘a proper theory of indexicals should be able to provide a unified account of anaphoric and non-anaphoric uses’ (2002, p. 118). With this goal in mind, I will follow Heim and Kratzer’s approach of assuming that hearers follow a general strategy when interpreting all ‘referring pronouns’:

In disambiguating the pronoun’s reference, listeners assign it to the most salient individual that allows them to make sense of the utterance (p.240).

I will reinterpret this general strategy in terms of a relevance theoretic account which distinguishes the salience of objects from the accessibility of mental representations of those objects, and in which the use of a referring pronoun is one option available to a speaker aiming at optimal relevance.

6.3.3 Pronouns: procedural and truth-conditional

In chapter 4, I touched on relevance theory-based approaches to pronouns (e.g. Wilson & Sperber, 1993; Hedley, 2005), which treat pronouns as encoding procedural information that contributes to the truth-conditional content of an utterance. In this section, I revisit these accounts, and consider what this procedural information might look like, and how it might interact with the discourse context.

Wilson and Sperber (1993) treat pronouns as encoding procedural constraints on explications. Following Kaplan (1989), they reject the view that the first person singular pronoun ‘I’ encodes the concept ‘the speaker’. On this analysis, an utterance of (12) would express the proposition in (13), and would therefore be necessarily false.

(12) I do not exist.

(13) The speaker of (12) does not exist.

Whilst (12) will be false whenever it is uttered, we do not want to say that it is *necessarily* false. Instead, according to Wilson and Sperber, the pronoun ‘I’ encodes procedural information which instructs the hearer to ‘identify its referent by first identifying the speaker’ (1993, p. 22). It is not the pronoun itself that enters into the proposition expressed, but a concept that picks out the referent of the pronoun on that particular occasion: pronouns ‘guide the search for the intended referent, which is part of the proposition expressed’ (1993, p. 23). For instance, when (12) is uttered by David Kaplan, it expresses the proposition in (14).

(14) David Kaplan does not exist.

Hedley (2005) outlines further arguments for treating pronouns as both procedural and truth-conditional. As he puts it, pronouns look ‘distinctly procedural’. Whilst ‘we clearly have a “concept” HAWK...we don’t seem to have a concept “HE”’ (2005, p. 42). Instead, the meaning of ‘he’ is context sensitive and in some sense means whatever it refers to in a particular utterance. Furthermore, in order to judge the truth or falsity of an utterance, reference must be assigned to any pronouns it contains.

(15) He has a big hat.

We cannot evaluate the truth of (15) unless we recognise the intended referent of ‘he’.

Assuming that pronouns are procedural, then, what sort of procedure might they encode?

6.3.4 A pronominal procedure

Hedley (2005) proposes an account of pronominals which is broadly compatible with the relevance theory approach to referring expressions outlined in this thesis. On Hedley’s account, the procedure encoded by ‘I’, might look something like, ‘find an individual concept of the speaker’ (2005, p. 6), while for singular ‘you’, it might be ‘something along the lines of “find an individual concept of the hearer”’ (2005, p. 8). For Hedley, a pronoun is ‘a linguistic device that can be used by a speaker in order to aid the process of reference resolution for his hearer, while not expending too much time and energy in producing fully specific (and probably repetitive) descriptive noun phrases’ (2005, p. 7).

Although this is largely compatible with my approach to referring expressions, there are two respects in which I think Hedley’s account does not go far enough, both of which relate to the hearer’s use of the relevance theory comprehension procedure. First, according to Hedley, pronouns ‘achieve relevance by uniquely picking out an accessible individual from the context’ (2005, p. 3), and when processing a pronoun the hearer will look around for a ‘reasonably accessible referent’ (p.7). Bearing in mind the content of the relevance theoretic comprehension procedure, I suggest that we should go further than this. As I have argued, a hearer following the procedure will always test interpretations in order of accessibility, and will therefore look for the *most* accessible referent that seems likely to make the utterance relevant in the expected way. If this yields an overall interpretation which satisfies his expectations of relevance, then he will stop. Thus, the individual picked out by the pronoun should not merely be ‘reasonably’ accessible: it should be the most accessible one which leads to a satisfactory overall interpretation. As with the referring expression types discussed in chapters 4 and 5, the speaker may use information encoded by the forms to constrain the set of potential referents to a point where this is the case. In this section, I will take a closer look at pronouns to consider what sort of information this might be.

Second, if we take seriously the notion of optimal relevance and the claim that the hearer follows the relevance theoretic comprehension procedure, then the motivation Hedley suggests for a speaker’s choosing to use a pronoun (‘not expending too much time and energy in producing fully specific (and probably repetitive) descriptive noun

phrases') again appears too vague. In the first place, as noted with respect to Ariel's analysis (see footnote 1, this chapter), the crucial type of effort involved in achieving optimal relevance is hearer's processing effort rather than speaker's production effort⁴. The effort involved in processing longer or more complicated referring expressions should be assessed relative to the cognitive effects produced by alternative utterances with the same interpretation. Using a longer, informationally richer term would not simply be a more effortful way of achieving the same effects. Rather, the utterance conveys a presumption of its own (optimal) relevance. Thus, if a speaker uses, say, a definite description where a pronoun would have sufficed, the hearer is entitled to look for extra effects to justify the extra effort to which he has been put. As discussed in the previous chapters, there are cases where use of an informationally richer (in Ariel's terms, lower Accessibility) form than necessary leads to extra or different effects. So, use of a pronoun as opposed to some other form of referring expression is not simply an effort saving device. Rather it is one means by which the speaker may make her utterance optimally relevant. According to relevance theoretic principles, a speaker should use a pronoun when it involves the minimum of linguistic material needed to achieve the intended overall interpretation in what, for the hearer, will be the most economical way.

However, a survey of the data reveals that although the overall role played by pronouns (and other referring expressions) in achieving relevance remains the same, pronouns do not all function in the same way. As discussed in section 5.2.1.5, Kaplan distinguishes pure indexicals from demonstrative uses. Whereas the pronoun 'I' is classed as a pure indexical since it does not require any sort of accompanying demonstration, the third person pronouns 'she', 'he' and 'it' may be used demonstratively to pick out a certain referent. Certain pronouns also carry specific information relating to gender and number, and I consider how we might characterise this information, and what role it might play in reference resolution, by turning my attention to the third person pronouns in English.

⁴ See Carston (2005) and (Wilson & Wharton, 2006) for a discussion of hearer's effort and speaker's effort in relation to Grice and relevance theory.

6.3.5 Third person pronouns

6.3.5.1 Gender and procedures

Hedley suggests that the procedure encoded by the masculine third person pronoun in English is something like, ‘find an individual concept with the feature ‘male’ (2005, p, 9). In my terms, this amounts to suggesting that use of the pronoun of ‘he’ constrains the set of potential referents to just those that carry the feature ‘male’. Thus, consider the examples in (16) and (17).

(16) Pierre said to Natasha that he should go.

(17) Pierre said to Natasha that she should go.

The only difference between (16) and (17) is the gender of the third person pronoun, and yet the interpretations are very different. In (16), Pierre is telling Natasha that it is Pierre (or some highly salient other male) who should go, whereas in (17), Pierre is telling Natasha that Natasha (or some other highly salient female) should go. The sets of potential referents from which the hearer will resolve reference are mutually exclusive in these two examples. To follow Hedley’s analysis, in (16) it is the set of individuals with the feature ‘male’ and in (17) the set of individuals with the feature ‘female’ from which the referent should be drawn. In this section, I will explore the role of gender features in more detail, and argue that we should take a slightly different approach to the treatment of gender in these examples.

On Hedley’s account, it is natural to see the gender information encoded by the pronouns ‘he’ and ‘she’ as conceptual in nature. The properties of being male or female are, after all, fairly concrete, and could be represented by the concepts MALE or FEMALE in a Fodorian conceptual representation system, or ‘language of thought’ (Fodor, 1975; 2008). This would in turn suggest that the encoded meanings of pronouns should be seen as at least partly conceptual in nature. However, as Hedley argues, it is not the concept MALE that enters into the explicature of the utterance, but a representation of the referent itself. As evidence for this claim, he cites Powell’s (2002) discussion of the case of Dr James Barry. Dr Barry was an eminent physician who died in 1865 and was discovered, after death, to have been a woman. Compare utterances (18) and (19) as examples of how this situation might be described.

(18) When he was laid out after he died, they discovered that he was actually a woman.

(19) ?When she was laid out after she died, they discovered that she was actually a woman.

If the hearer of these utterances believes that Dr James Barry is a male⁵, then use of the masculine pronoun is the most efficient way to ensure that reference is resolved correctly. Using the feminine pronoun would constrain the set of potential referents to just those classed as female, and at the stage in the utterance at which reference resolution should take place, Dr James Barry would not (as far as the hearer is concerned) be amongst that set. Thus, Hedley treats the pronoun ‘he’ as encoding a procedure that instructs the hearer to resolve reference on an individual concept with the feature ‘male’. He specifies that this feature is ‘pragmatic in nature (in English) rather than being categorical or semantic’ (Hedley, 2005, pp. 9, fn14), and later elaborates on this:

Procedural indications like ‘male’, ‘speaker’, ‘plural’ etc in English all operate at the level of evidence provided by the speaker for his hearer... rather than as semantically constraining factors.(Hedley, 2007, p. 91).

Thus, Hedley moves away from an analysis where pronouns themselves encode specific conceptual information about inherent features of the referent. Rather, he treats pronouns as encoding pro-concepts: placeholders for concepts which must then be worked out contextually in determining the explicit content of the utterance.

Whilst I will follow Hedley in holding that the work done by pronouns is closer to being procedural than conceptual, I want to question his treatment of the encoding of gender features. Rather than analysing the gendered pronouns (and pronouns in general) as encoding features of the type MALE or FEMALE, which figure as concepts in the language of thought, I want to argue that they should be seen as encoding procedures that operate at a sub-personal level, without invoking personal-level concepts such as MALE or FEMALE at all.⁶ Before considering in more detail the role that

⁵ Or that the name ‘James Barry’ would generally refer to a male.

⁶ Here I follow the personal / sub-personal distinction first introduced by Dennett (1969). Sub-personal explanations of behaviour deal in ‘entities and properties that can be shown to play a causal role in the action or behaviour, without necessarily standing in rational or normative relations to it’ (Carston, 2002b, p. 131). See Elton (2000) for an introduction to the distinction.

such procedures might play in reference resolution, I will briefly consider how recent approaches to pronominal reference have treated gender, number and person features.

Adger & Harbour (2008) discuss various syntactic, semantic and morphological approaches to what they refer to as ϕ -features (phi-features), which include features relating to person, number and gender, as well as those associated with honorification and definiteness (see chapter 1, section 1.6.3). A standard account of these features, first proposed by Heim and Kratzer (1998), treats them as carrying presuppositions. On this analysis, referring pronouns are seen as encoding free variables whose interpretation is determined by the ‘physical and psychological circumstances that prevail’ when they are processed (1998, p. 234). The ϕ -features are ‘syntactically adjoined to pronominals and their...semantic contribution is a presupposition that restricts the range of the assignment of values to [the] variables’ (Adger & Harbour, 2008, p. 14). According to this account, if a speaker uses a masculine pronoun to refer to a female, she has not uttered something false, but merely something potentially infelicitous. Moreover, given the appropriate set of beliefs, the speaker can felicitously refer to a female (Dr James Barry) as ‘he’, since by doing so she is expressing the presupposition that Dr Barry is male. As Heim & Kratzer explain (for a scenario in which the gender roles are the reverse of those in (18)):

This is intuitively right. If the discourse participants mistakenly believe a male referent to be female, or if they are willing to pretend that they do, then indeed an occurrence of ‘she’ can refer to a man, without any violation of principles of grammar. (Heim & Kratzer, 1998, p. 245).

Although the speaker of (18) knows that Dr Barry was a woman, the purpose of her utterance is to draw attention to a contrast between the actual state of affairs (Dr Barry being female) and the supposed state of affairs at the time described (Dr Barry being male). Use of an utterance with a false presupposition is therefore appropriate. On this account, the presupposition may be seen as a contextual assumption which forms part of the speaker’s and hearer’s common ground, and it would therefore have to be formulated in conceptual terms.

Adger & Harbour go on to consider an important issue relating to gender, and to ϕ -features in general: ‘gender is of two-types: semantically contentful and purely grammatical’ (2008, p. 13). Evidence from German seems to suggest that gender agreement is purely syntactic. For instance, the noun ‘Mädchen’ (‘girl’) is grammatically neuter, and this affects the form of the relative pronoun in (20):

(20) Das Mädchen, das /*die das Buch ließt.
 The.NEUT girl that. NEUT/that.FEM the book read.PRES

‘the girl who reads the book.’

Here, the feminine relative pronoun would be ungrammatical, which suggests that agreement is sensitive to syntactic rather than semantic features. However, the case of collective nouns in British English and certain instances of deferred reference complicate the picture. As noted by Adger & Harbour, both (21) and (22) are grammatical in some dialects of English.

(21) The committee has voted today.

(22) The committee have voted today.

The agreement on the auxiliary verb depends on the speaker’s intention to pick out either a collective action or a series of individual actions, rather than on syntactic number features associated with the noun. The same pattern can be found with a wide range of nouns denoting groups.⁷ Similarly, consider Adger & Harbour’s example of deferred reference in (23) (adapted from Nunberg’s (1979) ‘ham sandwich’ example):

(23) The hash browns at table six is / *are getting angry.

Here, the noun phrase ‘hash browns’, although grammatically plural, is used to refer to a customer in a restaurant, and on this interpretation the agreement on the verb must be singular. It has been argued (e.g. by Dowty & Jacobson, 1988), that such examples are problematic for analyses which treat agreement as purely syntactic.

I want to argue that the gender information encoded by the third-person singular pronouns in English lacks conceptual content, and should therefore be seen as more grammatical than semantic. Although, the grammatical gender of the pronoun often coincides with what we might think of as the ‘natural’ gender of the referent, it does not

⁷ Sauerland & Elbourne (2002, p. 289) provide the following (incomplete) list of examples: ‘cabinet’, ‘committee’, ‘platoon’, (political) ‘party’, ‘pride’, ‘hive’, ‘team’, ‘regiment’, ‘battalion’, ‘bank’, ‘government’, ‘group’, ‘family’, ‘faculty’, ‘Senate’, ‘House’ (of Lords, Commons, Representatives), ‘set’, ‘squad’.

follow that the procedures encoded by gendered terms draw directly on concepts such as MALE or FEMALE. Rather, they may be driven by the vestiges of a grammatical gender system, and indeed, we see similar examples with grammatical number features. For instance, there are pairs of nouns which, whilst very similar in terms of their conceptual content, vary as to their grammatical number features. Consider the examples in (24)-(27).

(24) The oats are rotten.

(25) The wheat is rotten.

(26) The woods are dark.

(27) The forest is dark.

The pairings of oats/wheat and woods/forest are roughly semantically equivalent (or at least may be used by some native speakers to refer to the same entities), and yet they differ in grammatical number. Similarly, words like ‘trousers’ and ‘scissors’, which pick out a single item, are grammatically plural⁸. The grammatical number feature associated with a certain noun phrase may therefore be distinct from any semantic or natural number associated with the encoded concept.

Given this assumption, how might we account for the collective noun and deferred references cases in (21)-(23)? Sauerland & Elbourne (2002) offer a tentative solution. They note that collective nouns behave as if they were ‘simultaneously singular and plural’ (p. 290). For example, in (28) the noun ‘set’ takes a singular determiner, but a plural verb.

(28) This set are all odd.

They therefore suggest that we are dealing not with one but with two categories of features, both of which take the values singular and plural. First, there is the standard number feature which indicates how many things are being referred to. In (28), there is a single set, and so the determiner must be singular. (29) is ungrammatical.

⁸ I suspect such examples may be a red herring. In each case they are objects which normally/usually come in pairs. When prefixed by ‘a pair’ they are syntactically singular, but when this is omitted plural agreement is expected, as would be the case with ‘the pair of gloves is...’ and ‘the gloves are...’

(29) *These set are all odd.

In addition, there is a feature system that Sauerland & Elbourne term ‘mereology’, which ‘indicates whether or not the entity under discussion is being conceived of as consisting of more than one member’. In their view, verbal agreement can refer to either feature, but ‘determiner concord refers only to the Number feature’ (p. 291). This approach provides some insight into what is going on in utterances such as (28), despite the apparent clash in number feature between determiner and verb. However, the introduction of the mereology system does not completely counter the argument that collective noun cases are problematic for a purely syntactic analysis of number features. The speaker often has a choice about which type of feature (number or mereology) to use in verbal agreement, and that choice will be driven by subtleties in the interpretation that the speaker intends to communicate.⁹

I suggest that collective nouns are, by their nature, exceptional. As grammatically singular nouns that represent a set of individual items, they can be used to refer either to the set as a whole, or to all the individual members of the set. The choice between singular and plural agreement is therefore not free, but depends on the intended interpretation of the utterance. Consider (30).

(30) The committee were /?was all late for the meeting.

For use of the word ‘all’ to be relevant in this utterance, the hearer must represent the committee as a group of individuals. Thus, the speaker uttering (30) is understood as conveying that all members of the committee were late for the meeting. Although it is underlyingly singular, the interpretation of a collective noun ‘the X’ can always be pragmatically adjusted in context to convey ‘members of X’. In that case, they trigger plural agreement, which acts as a clue to the hearer that this is the intended interpretation.

Returning to the treatment of gender, I therefore propose that the gender information carried by third person pronouns in English is grammatical rather than

⁹ Deferred cases such as (23) (‘The hash browns is getting angry’) are seen as problematic for a purely syntactic account of number features. However, I suggest that these work differently from the collective noun cases. Whereas the mereology is part of the definition of collective nouns (a collection of individual parts making up a collective whole), this is not the case for deferred uses of non-collective nouns. Rather, it seems that in example (23) we have a case of meaning transfer. The NP ‘hash browns’ is used as the name of another entity to which it stands in a ‘salient functional relation’. See (Nunberg, 2004, p. 346) for further discussion of deferred reference. I will therefore leave these cases aside.

semantic or conceptual. The speaker's choice of a masculine pronoun in preference to a feminine pronoun (or vice versa) does not in general take place at a personal level, using concepts drawn from the vocabulary of the language of thought. It is not necessary for a speaker to check whether the intended referent falls under the concept MALE.¹⁰ Rather, a pronoun constrains the set of potential referents to just those belonging to a category of items that are identifiable in sub-personal rather than personal terms (perhaps using the notion of a 'phi-feature' construed as grammatical rather than semantic). The ability to identify items belonging to this same category must play a role not only in comprehension, but also in language production: it is just the category of items for which that particular pronoun can be legitimately used. On the approach I propose, then, the masculine pronoun 'he' encodes both a pro-concept and a procedure which restricts the set of potential referents to just those belonging to a category identifiable in sub-personal terms as containing items that can be legitimately referred to using the pronoun 'he'. On this sub-personal approach, speakers might be seen as having, as part of their knowledge of the word 'he', a detector for items which fall into this category. Typically, detectors operate on the basis of cues which pick out stereotypical or prototypical members of a category, but these cues do not amount to a definition.¹¹ Having restricted the set in this way, reference resolution proceeds, as described in the previous chapters, with potential referents from this sub-personally identifiable category being tested in order of accessibility.

This move away from semantic/conceptual gender and number features makes it easy to account for the Dr James Barry type cases, and also sheds interesting new light on other seemingly exceptional examples. Whilst we might make the broad generalisation that the 'he'/'she'/'it' distinction in English pronouns correlates with the distinction between human males, human females and inanimate objects respectively, there are many less clear-cut cases. A ship (or car, or country), for example, may be referred to as 'it' or 'she'. In my framework, this is simply because it is identifiable in sub-personal terms as belonging both to the set of things that can be referred to as 'it' and the set of things that can be referred to as 'she'.¹² A similar explanation can be given for the otherwise messy issue of how to refer to animals using a pronominal form.

¹⁰ See Corbett (1991) for a discussion of semantic versus natural gender.

¹¹ A parallel might be drawn with the visual systems of frogs. According to Letvin, Maturana, McCulloch and Pitts (1959), the frog's fly detector is based on the ability to pick out small black moving dots, and so does not pick out all and only flies. Also see Agar (1993) for further discussion.

¹² There may be individual differences in how acceptable different speakers find these two possibilities, and this further supports the sub-personal procedural analysis. If it is becoming less common to refer to ships as 'she' (rather than 'it'), we can see this as a change in the sub-personal categorization rather than as a result of ships being less feminine in any full-fledged conceptual sense.

A cat or dog may be referred to as ‘he’, ‘she’ or ‘it’ because it may belong to all three sub-personally identifiable sets, and not because of the natural gender (if known) of the animal itself. Membership of a certain set is therefore dependent on how the particular speaker categorises the particular animal (or how they assume their hearer categorises it), and whilst this may be influenced by natural gender, it is far from determined by it. Gender sets may overlap, and their membership may vary as the discourse context, including the speaker’s assumptions and communicative intentions, vary. A speaker may well refer to a cat as ‘she’, and then change to ‘he’ on discovering that the cat is male. This does not make the original utterance false, or the original act of referring necessarily infelicitous.

Although animals may belong to the set of things that can be referred to using the pronoun ‘it’, as in (31), this is not usually the case with humans, even if the gender of the referent is unknown, as in (32).

(31) I’m going to buy [a dog]_i and call it_i Rover.

(32) *I’ll give this to [whoever wins]_j and it_j can keep it.

An exception seems to be the case of unborn babies, where in certain circumstances ‘it’ is just as acceptable as ‘they’.

(33) If I ever have a baby, it’s going to have the best of everything.

(34) If I ever have a baby, they’re going to have the best of everything.

Until the natural gender of the baby is known, it will belong to the set of things that can be referred to using ‘it’, as in (33) or ‘they’, as in (34). Again we can understand this as a reflection of how the speaker categorises an unborn child, rather than of the child’s falling under one or other full-fledged concept. Furthermore, examples such as (34) provide evidence that the grammatical number feature on the plural pronoun does not necessarily correspond to semantic plurality. There is no suggestion in (34) that there is more than one baby.

I have argued that pronouns encode procedural rather than conceptual information. However, some recent relevance-theoretic accounts (Hussein, 2008; Zaki, 2009) propose a mixed analysis, in which pronouns are seen as encoding both

conceptual and procedural information. The idea behind these accounts is that pronouns (and some other forms of referring expression) encode procedures which guide the process of reference resolution by activating conceptual information about proximity, gender and number. Hussein (2008, p.72) goes so far as to claim that an utterance containing the pronoun ‘he’ ‘would be false if “he” is taken to refer to someone who is not “singular” or “masculine”’. In the previous chapter I argued against attempts to treat features such as ‘proximal’ and ‘distal’ as corresponding to full-fledged concepts or semantic notions¹³, and in this chapter I have tried to do the same with gender and number features. The information encoded by pronouns is no more than a clue to help guide the hearer to the intended meaning. On this account, the gender and number information associated with the pronoun does not itself enter into the proposition expressed, nor does it act as a presupposition, as proposed by Heim and Kratzer. Rather, the choice of pronoun guides the hearer as he resolves reference and it is a representation of the referent, rather than a full-fledged concept such as MALE or FEMALE, which enters into the explicitly communicated meaning. From now on, I will therefore treat pronouns as fully procedural in nature.

6.3.5.2 The role of contextual assumptions

My account of pronouns (and referring expressions in general) relies heavily on the relevance theoretic comprehension procedure. However, it is important to bear in mind that this procedure deals with the overall interpretation of an utterance and that the hearer must construct hypotheses relating to each of the three sub-parts (explicit content, contextual assumptions and contextual implications). In this framework, it is quite possible that the information encoded by a pronoun will not determine a unique referent as the most accessible one. Rather, a speaker aiming at optimal relevance should formulate her utterance so that the intended referent is the most accessible candidate which can combine with accessible contextual assumptions to yield an interpretation that is relevant in the expected way.

To see why this is important, consider first a simple example of how a procedural analysis of third person singular pronouns might go.

(35) John went to the shops. He bought a red jumper.

¹³ On my sub-personal account, a range of factors, including spatial, temporal, discourse and emotional factors, will go into the online computation of whether an item falls into the category of things that may be appropriately referred to as ‘this’ or as ‘that’. Hence, use of a demonstrative determiner is not just a matter of checking whether the item falls under a fully-fledged concept of NEAR or FAR.

In order to derive the explicit content of (35), the hearer must assign reference to the pronoun 'he'. Assuming that no extra-linguistic referents are available in the discourse context, it should be obvious to both speaker and hearer that the most accessible potential referent will be 'John'. This, tested in a highly accessible context which includes the contextual assumption that people go to the shops to buy things, yields an interpretation that might be relevant enough to be worth the hearer's attention in a way the speaker could manifestly have foreseen.

Since pronouns provide minimal constraints on the set of potential referents, my account predicts that they will be used when little or no constraint is needed. In example (35), there is only one candidate referent in the discourse context. However, consider example (36), where there are in principle two highly accessible potential referents ('Britney' and 'Justin') for the pronoun in the embedded clause.

(36) Britney told Justin that he couldn't sing.

I have argued that the pronoun 'he' constrains the set of potential referents to a sub-personally identifiable set of items that can be legitimately referred to by the pronoun 'he'. Human beings with masculine names will generally belong to this set, whereas human beings with feminine names will not. On normal encyclopaedic assumptions about the names 'Britney' and 'Justin', there will therefore be only one candidate referent that fits with this constraint, and so reference will be resolved on 'Justin'.¹⁴

Now compare (36) with (37) and (38) (adapted from Nieuwland & Van Berkum (2006)), ignoring the possibility of non-anaphoric readings:

(37) Britney told Justin that she couldn't sing.

(38) Britney told Christina that she couldn't sing.

Without any details of the context in which (38) is uttered, it is unclear whether the pronoun 'she' is intended to refer to Britney or Christina. Although in (36) and (37) the procedure encoded by the pronoun sufficiently constrains the set of possible referents to a point where there is only a single highly accessible referent, this is not the case in

¹⁴ It is of course logically possible that 'he' is intended to refer to another male entity; but if this were the intended interpretation, we would expect the speaker to narrow the context for resolution using extra-linguistic means such as gesturing, and/or to demand more effort from the hearer by stressing the pronoun, and thus encouraging them to look for more or different effects.

(38). Both ‘Britney’ and ‘Christina’ will belong to the sub-personally identifiable set of things that can be referred to by the feminine pronoun. The procedural constraint encoded by the pronoun leaves more than one possible referent. However, reference does not necessarily have to be resolved at this point for the utterance to be optimally relevant. According to the relevance theoretic comprehension procedure, hearers test ‘interpretive hypotheses’, not just potential referents. Reference will eventually be resolved as a result of combining a possible referent with a possible context. Thus, the speaker need not always constrain the set of potential referents to a single one, as long as there is a highly accessible context with which one can combine to yield a relevant overall interpretation, while others do not. The effect of context selection on reference resolution is therefore crucial, and is missing from the previous accounts (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993; Reboul, 1997).

Continuing with the analysis of (38), we can see that choices of context will lead to different potential interpretations¹⁵. Suppose that in context 1, the speaker is describing an argument that she has just witnessed between Britney and Christina, while in context 2, the speaker is describing a whispered conversation that she has just overheard between Britney and Christina. The contextual assumptions that are likely to be highly accessible to the hearer will be different in these two contexts. Context 1 may give easy access to assumptions such as (39) and (40).

(39) When arguing, people say insulting things to one another.

(40) Telling someone that they cannot sing is an insult.

Interpretation of ‘she’ as referring to Christina is therefore likely to be preferred in this context, since it can combine with these highly accessible contextual assumptions to make the utterance relevant as expected.

These examples illustrate some of the ways in which third person pronouns can interact with the discourse context. In other examples, however, it is not so clear that the encoded procedure contributes directly to the reference resolution process at all. In the next sub-section, I will look at some of these.

¹⁵ See Nieuwland & Van Berkum, (2006) for neurolinguistic evidence of context influencing pronoun interpretation in this way.

6.3.5.3 Gender-redundant uses

Consider utterance (41).

(41) I spoke to Andrey and he said that all was well.

If we apply the reference resolution process outlined above, then any specific information associated with the masculine pronoun appears to be redundant in this case. In the discourse context, there are two explicitly mentioned, and therefore highly accessible, potential referents: the speaker of the utterance and Andrey. Removing the pronoun entirely yields a reading where the speaker is the understood subject of the verb 'said', as in (42).

(42) I spoke to Andrey and said that all was well.

By contrast, utterance (41) leads to an interpretation where reference is resolved on Andrey. It therefore seems reasonable to assume that the speaker is the most accessible potential referent in the discourse context, and that use of the pronoun in (41) eliminates the speaker as a potential referent, leaving Andrey as the first to be tested. Thus, the use of a third person pronoun narrows the hypothesis space for the inferential process of reference resolution.

On this account, the fact that Andrey belongs to the set of things that can be appropriately referred to by 'he' (as opposed to 'she') is irrelevant for purposes of reference resolution. The grammar of English forces the speaker to use an appropriately gendered version of the third-person pronoun when referring to humans. However, what is significant is not the encoded gender information but the fact that the third person pronoun has been used, since this leaves the intended referent as the most accessible one that meets the constraints imposed by 'he'. Although a representation of the speaker might be generally more accessible than one of Andrey at this point in the utterance, the speaker (and likewise the hearer) would not normally be a member of the set of things that can be appropriately referred to in the third person. This rules out an interpretation where reference is resolved on the speaker, and the hearer will continue to test possible interpretations in order of accessibility. Andrey is the next most accessible interpretation, and he belongs to the set of things that can appropriately be referred to using the pronoun 'he'. If the resulting interpretation is relevant in the expected way, the hearer will accept it as the intended meaning.

As with the gender information discussed above, on this approach, features such as ‘first person’, ‘second person’ or ‘third person’ are seen as purely grammatical. The use of a third person pronoun does not carry conceptual information about the intended referent but rather indicates that this referent belongs to a category identifiable on a sub-personal level. This suggests an explanation for why it is sometimes possible for a speaker to refer to herself (or her hearer) using the third person. In certain discourse contexts, the speaker herself will belong to the sub-personally identifiable set of things that can be appropriately referred to using a third person pronoun.

6.3.5.4 Informative uses

Whilst the procedural information provided by the pronoun is crucial to reference resolution in cases such as (16) and (17) above (‘Pierre said to Natasha that he/she should go’), on other occasions the speaker may use the choice of pronoun in an informative way, to contribute to other aspects of the representation that she is encouraging her hearer to build. Consider (43).

(43) I met a friend for lunch and she had the fish.

In interpreting the first part of (43), the hearer has to construct a representation of the speaker’s friend. Since ‘friend’ is gender-neutral, he may assume that the intended relevance of the utterance does not depend on the gender of this friend. However, in the second part of the utterance, the speaker uses a gendered pronoun, and the hearer must therefore narrow the interpretation of ‘friend’ to include only things that can be appropriately referred to by the pronoun ‘she’. As discussed above, a speaker is constrained by the grammar of English to choose the appropriately gendered pronoun for the intended referent in such cases. Moreover, if the hearer is resolving reference on a definite person, knowing the gender of that person is likely to increase the relevance of the utterance by giving him access to additional implications (or other cognitive effects) at no extra cost.

Recall, however, that a speaker aiming at optimal relevance need only make her utterance as relevant as is compatible with her own preferences and abilities. There may be situations in which the speaker does not know, or does not wish to disclose, some information about the gender of the referent that might indeed increase the relevance of her utterance. In such cases, the gender neutral pronoun ‘they’ may be

substituted for 'he' or 'she' in some dialects. Thus, imagine (44) spoken by a wife to her suspicious husband when she arrives home late.

(44) I met a friend for a drink and they talked a lot!¹⁶

Here, the speaker does not wish to lie, but also does not wish to disclose any further information about the gender of the friend. However, such obvious avoidance of a gendered pronoun might well lead the hearer to access certain contextual assumptions about what information wives might wish to withhold from their husbands, and thus infer that the friend was a male. The fact that utterance (44) could carry this implication suggests that even when irrelevant for the purposes of reference resolution, it is normal - and hence less costly for both speaker and hearer in terms of processing effort - to select a gendered pronoun when referring to a human being whose gender is known. Choosing not to do so may imply (or implicate) something about the speaker's abilities or preferences.

Whereas in (44) it is the speaker's preferences that drive her choice of pronoun, in (45), as with the unborn baby example in (34), her choice is restricted by her abilities.

(45) I'll give this to whoever wins and they can keep it.

Here, the speaker does not know whether the winner will belong to the set of things that can be appropriately referred to by 'he' or the set of things that can be appropriately referred to by 'she', and so she uses the grammatically plural but gender neutral pronoun 'they'.

6.3.6 The first person singular pronoun

The majority of null subject utterances in the diary style registers are interpreted as first person singular, and for this reason I turn my attention again to the procedural information encoded by the pronoun 'I'. In many ways, analysis of this first person

¹⁶ The fact that in such uses the verb agrees with the (plural) number feature on the pronoun is further evidence that such features are syntactic in nature.

(i) A friend recommended this. They swear by it.

(ii) * A friend recommended this. They swears by it.

case may seem more straightforward than for the third person pronouns. We are not so easily distracted by the superficially conceptual notion of gender (and to a lesser degree number). As discussed in section 6.3.2, it soon becomes clear that defining ‘I’ as meaning ‘the speaker of the utterance’ leads to incorrect predictions about the truth-values of utterances in crucial cases. Rather, for Wilson and Sperber, ‘I’ is seen as instructing the hearer to ‘identify its referent by first identifying the speaker’ (1993, p. 22), and for Hedley (2005, p. 7) it is seen as encoding a procedure which tells the hearer to find ‘an individual concept of the speaker that is relevant in the particular context’.

Turning to my account of referring expressions, and following the discussion of third person pronouns above, the procedure encoded by ‘I’ will add an extra layer of activation to the sub-personally identifiable set of items that can be appropriately referred to (in that situation) using the 1st person singular pronoun. Obviously, this set varies from speaker to speaker, and in most cases will contain just the speaker herself¹⁷.

To shed further light on the procedure associated with the first-person singular pronoun, I will next consider the much discussed examples of answering machine and post-it-note utterances. The so-called answering machine paradox (Sidelle, 1991; Predelli, 1998; Corazza, Fish, & Gorvett, 2002; Powell, 2002; Hedley, 2005) involves utterances such as (46), recorded on answering machines or left as notes.

(46) I am not here now.

Compare this utterance with (47), which, according to Kaplan, is ‘deeply, and in some sense...universally true...it cannot be uttered falsely’ (1989, p. 509).

(47) I am here now.

Following Kaplan’s logic, (46) should be universally false whenever it is uttered. However, this goes against the strong intuition that tokens of (46) recorded on answering machines or posted on office doors may be both informative and true.

Corazza et al (2002) propose an analysis designed to deal with such examples, which they call ‘the conventional account’ (2002, p. 12). On this account, speakers and

¹⁷ However, this analysis might help to explain the sort of deferred reference cases discussed by Powell (2002, p. 137): for example, (iii) as uttered by a race-goer who has placed a bet on a particular horse.

(iii) I am in last place

Such examples suggest that not only the speaker, but items that stand in an appropriate relation to the speaker, may belong in the category of items picked out by the use of ‘I’.

hearers are aware of certain social conventions relating to the use of answering machines, post-it notes, etc, and it is because of this awareness that they can felicitously produce and understand utterances such as (46). According to Corazza et al, ‘without such settings and conventions, we would be unable to successfully use and manipulate answering machines and other similar devices’ (2002, p. 13). I find this appeal to convention rather unsatisfactory. Whilst I agree that the discourse context is crucial in determining how an utterance will be understood, Corazza et al suggest no explanation for how the interpretation of certain utterances in certain discourse contexts should have become conventionalised in this way. Their account raises the question of how the first users of answering machines dealt with such utterances. I suspect that early individual users of answering machines would have had no problems communicating and interpreting the intended meaning, even though the devices were novel and innovative.

An added variation (Sidelle, 1991, p. 5) complicates the story still further. In this version, utterance (46) (‘I am not here now’) is not left on an answering machine, but posted on the office door of an absent university professor. However, it was not written by the professor herself, but by a university administrator who is aware of the professor’s absence, and who has noticed a stream of confused students knocking on her door. Although the professor is the intended referent of the pronoun, she did not produce the utterance. The administrator may have been asked by the professor to write the note on her behalf, in which case she is, in effect, quoting the professor. Alternatively, the professor may have no knowledge of the note, and so the administrator’s act becomes genuinely deceptive, in that she is purporting to speak on the professor’s behalf when she is not entitled to do so.

A relevance-based account of referring expressions and utterance interpretation in which contextual assumptions play a central role eliminates the need to appeal to conventionalisation, and offers a new perspective on these cases. Placing the note on the door is an act of ostensive communication, which creates a presumption of optimal relevance. The audience will follow the path of least effort when testing possible interpretations, and the writer will construct her utterance so that this strategy will lead the hearer to the intended interpretation without putting him to any gratuitous effort. Given the discourse context, the administrator knows, on the one hand, that the audience will be unaware of who actually wrote the message, and, on the other hand, that the most accessible referent for them will be the professor. Moreover, the administrator is aware that a scenario in which the owner of an office puts a note on their own door relating to their own whereabouts will be highly accessible to the

audience. Given this highly accessible contextual scenario, the first interpretation the audience is likely to test will involve the pronoun ‘I’ being taken to refer to the professor. Since this interpretation will be relevant in the expected way, the audience will consider no other possibilities, and will settle on this interpretation¹⁸. Thus, the interlocutors need not rely on conventions, but rather on highly accessible assumptions about how humans are likely to behave in certain situations. Even if a student were to see the administrator writing and attaching the note, the assumption that she is doing this on behalf of the professor is likely to be much more accessible than a scenario in which the administrator attaches a note to someone else’s office to inform students of her own absence. In that case, the administrator will be seen as quoting (or purporting to quote) the professor, and the utterance will be understood as such. As a sophisticated understander (Sperber, 1994), the student should be able to recognise that the administrator is taking the perspective of the professor, and will therefore look for an interpretation that the speaker (in this case the administrator) ‘might have thought he would think was relevant enough’ (Wilson, 2000, p. 422). An interpretation where the administrator is taking the perspective of the professor would then be relevant in the expected way.

Examples of this type are not necessarily restricted to written notes or recorded messages. Consider an utterance of (48) in the following discourse context. Clare is at her desk in her place of work. However, she is not officially at work, but rather studying for her evening course during an afternoon of annual leave. A colleague approaches her with a work-related query and she utters (48).

(48) I’m not here!

Although it seems that Clare has said something obviously false, the hearer is entitled to assume that her utterance is optimally relevant. In this case I suggest that the proposition expressed is along the lines of (49), where ‘I’ functions in the normal way, and [IN THE OFFICE]* is an ad hoc concept expressing something like the narrowed meaning: ‘in the office for the purposes of performing my usual duties of employment’.

¹⁸ The administrator could, of course, have achieved a similar outcome by using a non-first person referring expression (iv).

(iv) Professor Smith is not here today

This might, however, lead the hearers to derive certain implications that the administrator might wish to avoid. Perhaps she does not want the students to form the opinion that the professor was careless not to leave such a note herself, to realise that the administrator has further information, and so come and pester her.

(49) Clare is not [in the office]*

To sum up, I propose that the first person singular pronoun encodes a procedure which, in effect, restricts the set of potential referents to just those that the speaker can refer to using the term 'I' in the discourse context. An important advantage of such an analysis is that it suggests a possible explanation for the less straightforward cases. In the case of the note on the door in (46), the discourse context is such that the professor can be referred to by use of 'I' even though the utterance was produced by someone else. In cases of deferred reference, something associated with the speaker can belong to the set of things to which 'I' can be used to refer. Again, such examples persuade us away from a 'feature' analysis in which the pronoun is seen as encoding semantic or conceptual information. Rather, the procedures associated with pronouns operate at a sub-personal level.

6.4 Relevance theory and null subjects

In this section, I will argue that the occurrences of null subjects in non-*pro*-drop languages are not surprising on a relevance-theoretic approach to utterance production and interpretation. Rather, they should be expected to arise spontaneously under certain circumstances as speakers aim at optimal relevance. I will divide the null subjects into three categories, and argue that the driving force underlying each category is the balance between hearer's effort and cognitive effects. Ultimately, on my account, the diary style nulls emerge as simply another type of referring expression available to a speaker, which will interact with the discourse context to yield an interpretation that makes the utterance relevant in the expected way.

6.4.1 Greater or different effects

Following the relevance-based approach outlined in chapters 4 and 5, it could be tempting to claim simply that in contexts such as diaries and email messages, the (usually first person) referent is so accessible that the procedural information encoded by the pronoun is no longer needed. Although I will argue that this is indeed the case, it does not on its own fully explain the distribution of null subjects across language use in English. If the accessibility of referents were the only significant factor, then we might expect to see null subjects in many more discourse contexts than we in fact do. We might predict that first person null subjects would occur spontaneously whenever the

speaker is the most accessible potential referent. However, this is not the case. Rather, we find that null subjects are restricted for the most part to informal written (and sometimes spoken) registers. The fact that null subjects in English are on the whole associated with casual uses of language will be crucial to my analysis of them as relevance-driven referring expressions.

In section 6.3.5.3, I showed how procedural information associated with the gender of a pronoun may in some cases be redundant, and provided evidence that, despite this redundancy, the appropriately gendered pronoun is selected. Outside diary contexts, speakers do not normally omit a gendered pronoun (or substitute it with a neutral one) simply because the information is not required. In general, the grammar of English requires arguments to be phonologically realised, so even when the intended referent is already the most accessible one in the discourse context, an appropriate pronoun must be used.

This brings me back to the question I posed at the end of chapter 1: what are the contextual and processing conditions under which this usually strict requirement that subjects be overtly articulated might be relaxed in certain registers of non-*pro*-drop languages such as English? In this section, I will attempt to answer this question by arguing that the restrictions may be relaxed in several different sets of circumstances, each of which ultimately comes down to the balancing of effects for effort. Thus, the use and distribution of English null subjects is ultimately driven by the speaker's aim of achieving optimal relevance.

6.4.2 Greater effects and different effects

6.4.2.1 The informal null

The informal (usually written) registers in which many instances of English null subjects occur tend to have the speaker's presence explicitly acknowledged elsewhere in the text. Diaries are essentially first person narratives in which the author and her activities are by definition the subject matter. Emails and SMS texts are direct messages from one individual to another (or others), and include 'from' and 'to' fields explicitly stating who the speaker is. Of course, not all diaries, emails or text messages utilise null subjects. As with spoken English, it seems to be the casual, informal nature of the texts that triggers use of the null subject option.

Although, as we have seen, there may be many instances in which the procedural information encoded by pronouns is superfluous, by dropping this normally required

information, the speaker overtly indicates that this is indeed the case. As a result, she communicates that the mutual cognitive environment she shares with the hearer is such that she can safely relax the normal requirements on formal language use, and still be confident that her message will be understood as she intends.

Sperber and Wilson (1986/95) argue that the style adopted by the speaker may communicate how she regards the relationship between herself and her hearer or hearers:

A speaker aiming at optimal relevance will leave implicit everything her hearer can be trusted to supply with less effort than would be needed to process an explicit prompt. The more information she leaves implicit, the greater the degree of mutual understanding she makes it manifest that she takes to exist between her and her hearer (Sperber & Wilson, 1986/95, p. 218).

The use of null subjects (and other omissions and abbreviations) in these informal registers communicates that the speaker not only regards the intended referent as the most accessible one for the hearer, but is also quite confident in this respect. As Sperber and Wilson note, there is 'no entirely neutral style' (1986/95, p. 218). The speaker in the diary null contexts reveals that she assumes a 'certain degree of mutuality' between herself and the hearer, which makes it possible to trust the hearer to infer much of what she wishes to communicate without any overt linguistic prompt. As a result, the use of null subjects contributes to the overall tone of the discourse as relaxed, casual and intimate.

Although the addressee of a null subject utterance has less linguistic material to process, since null subjects are ungrammatical in most varieties of English, this option is in some sense marked. Given that marked options in general cost the hearer more processing effort, it seems reasonable to assume that subjectless utterances in English - at least when first encountered in a text - require more processing effort than their overt subject counterparts. However, this marked use and the implications that arise from it lead to extra effects in terms of the tone and style of the discourse, and justify the effort involved. If null subjects were unmarked in English, as in standard *pro*-drop languages, then we would be unable to explain why the null subject examples correlate

with the informal discourse contexts.¹⁹ It is the very fact that null subject sentences are marked in English that prompts the hearer to derive the extra effects relating to the tone and style of the discourse. Formal emails, diaries and text messages exist alongside the casual examples in which we find the null subjects. This is to be expected if we accept that there will be cases where, even though the identity of the referent might still be highly accessible, the speaker does not wish to create the same degree of intimacy. Compare (50) and (51).

(50) I am pleased to confirm that the Grant of Probate has now been issued.

(51) Forgot to mention, saw my ex-boyfriend Mike at the weekend.

Both are from personal emails, and in both cases the intended referent of the subject referring expression is the most highly accessible given the discourse context²⁰. However, the discourse context and the nature of the relationship between speaker and hearer (solicitor and client in (50) as opposed to friends in (51)) means that it is appropriate to imply intimacy in (51), but not (50). The overtness of the pronoun does not change what is explicitly communicated in either case, but rather contributes to the style and tone of the discourse. Thus, it is not that informal registers licence the use of null subjects, but rather that the use of null subjects is one way in which the speaker can set an informal tone for the discourse.

Bridget Jones' Diary (Fielding, 1996) features two distinct writing styles. The 'diary style' sections which feature extensive use of dropped subjects and other missing elements are found alongside much more traditional narrative episodes. Whilst both are purportedly extracts from a diary, sections featuring diary-style omissions focus strongly on the narrator's private thoughts, feelings and emotions. The remaining sections describe events and incidents in a more detached and traditional first-person narrative style. Thus, the tone of the writing reflects the subject matter, and directly mirrors the level of intimacy with, and empathy for, the narrator that the reader is expected to feel.

¹⁹ This, in turn, predicts that overt subject use in *pro*-drop languages should create an extra effect of some sort. This seems to be the case. McCloskey & Hale (1983, p. 491) observe that in such languages, 'use of a pronoun subject is associated with emphatic or contrastive focus on the subject'.

²⁰ Further to this, the form of the copula 'am' in (50) makes it clear that the subject is the first person singular.

In this section, I have argued that some instances of diary null subjects are driven by the informality of the registers in which they occur. According to clause (b) of the definition of optimal relevance, the hearer is entitled to assume that the utterance the speaker has produced is the most relevant one compatible with her abilities and preferences. In these cases, the speaker had the ability to produce overt subjects, so the hearer must conclude that she chose not to. This in turn can be interpreted in two ways. As discussed, the speaker should not be putting the hearer to any gratuitous effort. So her choice to omit certain elements might imply that she did not judge the information encoded by those elements to be necessary given the mutual understanding between the interlocutors. Alternatively, the choice may be driven not by the speaker's abilities but by her preferences. It could be that the speaker preferred to save her own effort rather than go to the trouble of explicitly articulating or writing the omitted elements. This decision to favour her own interests over the hearer's might play a significant role in diaries and note-taking, where the writer is also the intended audience. If read by someone other than the writer, omissions of this type must be in some sense condoned as less than stylistically optimal from the audience's perspective. These various possibilities do not explain all occurrences of null subjects in non-*pro*-drop languages. To account for the remaining nulls, I again turn my attention to the second clause in the definition of optimal relevance.

6.4.2.2 The pressurized null

As has often been pointed out, many of the null subject utterances in non-*pro*-drop languages occur in discourse contexts where there is a restriction on time or space. These include newspaper headlines such as (52), certain internet message boards or forums such as (53), postcards such as (7), text messages such as (55), telegrams such as (56), "tweets", such as (57) and note-taking, as in (58).

(52) Falluja insurgents say not holding Hassan.

(53) Was delivered promptly.

(54) Visited the castle yesterday. Wish you were here.

(55) Must have missed one another.

- (56) Impossible attend first night. Will attend second night if you have one.
- (57) Just smashed my shin on a big wooden elephant (long story).
(<http://twitter.com/MitchBenn> accessed 13th January 2010)
- (58) Can be integrated. Needs investigation. (notes for minutes of a meeting to discuss updating a database system)

In such cases, either the length of the utterance is restricted in some way (even to the point where only a limited number of characters is allowed, as in some electronic formats), or the utterance is produced under pressure or time constraints. In relevance theory terms, the speaker's ability to produce a fuller version is in some way compromised. Recall that the speaker's overall aim is to produce an utterance which yields enough cognitive effects to justify the effort required of the hearer, and which will be the most relevant utterance compatible with the speaker's abilities and preferences (including any constraints on her production abilities). Relevance itself is 'a positive function of effects achieved, and a negative function of effort incurred' (Hall, 2004, p. 220). As discussed in chapter 4, procedural meaning contributes to the relevance of an utterance by guiding the hearer towards the intended effects, and thus reducing the effort required of him.

In these pressurized cases, the speaker only has a certain amount of space, time or characters at her disposal, and as a consequence something has to 'give'. In this section I consider why, in such cases, subjects are so often the thing that 'gives'. I have argued above that pronouns encode procedural meaning which contributes to the explicit content of an utterance. It is not, of course, the pronoun itself which enters into the explicit content, but (a representation of) the referent of the pronoun. So, it is not the presence of the pronoun itself which is significant in the null subject examples, but rather the procedural meaning it encodes, and the role this meaning plays in guiding the hearer to the intended referent in a particular discourse context. In section 6.3.5 above, I showed how the procedural information carried by third person pronouns may be either redundant or decisive in deriving the explicit content of an utterance. Bearing this in mind, I propose that in the pressurized cases, the chances that a speaker will drop a pronoun (or rather, that she will sacrifice the procedural meaning it encodes) will depend on what that meaning contributes to the explicit content of the utterance.

As discussed above, there are cases where it is very likely that the intended referent is already the most accessible candidate for the hearer in the discourse context. In such cases, the speaker is risking very little in terms of communicating the intended explicit content by omitting the pronoun. My approach predicts that first person pronouns will be omitted more frequently than others, since the speaker will very often be the most accessible potential referent in the discourse context. However, it does not rule out the possibility that other pronouns will be vulnerable, given a discourse context where the non-first person referent is the most accessible. My approach also predicts that expletive pronouns will frequently be candidates for omission, since they too contribute little to the explicit content of an utterance.

Given my treatment of pressurized null subjects within the overall account of referring expressions, it may be tempting to ask why omission of highly accessible subjects is not more widespread. Recall that according to relevance theory, communication takes place at a risk, and sending a hearer down the wrong inferential path (e.g. towards the wrong referent) is likely to be very costly in terms of effort. Processing a pronoun, on the other hand, is likely to be quite cheap in terms of hearer effort, both because they are high-frequency items and because of their limited semantic content. So, all else being equal, it does not make sense to risk misunderstanding for very little gain in terms of effort saved. In the pressurized discourse contexts, on the other hand, omission of the subjects frees up space, characters and/or time which the speaker can devote to encoding information which leads to cognitive effects, thereby making the utterance more relevant.

6.4.2.3 The ostensibly vague null

There are some instances of diary-style null subjects which appear to be neither driven by informality nor produced under pressurized circumstances. In these cases, the speaker may not know, or may not want to commit to, the exact identity of the referent. Such uses are discussed by Oh (2006, pp. 842-843), who notes that ‘zero anaphora can be deployed precisely in order to avoid the selection of a particular term and the stance that speaker would thereby be taking’. Alternatively, it may be that identification of an exact referent simply does not make much difference to overall relevance. I will call such instances the ostensibly vague null subjects. In these cases, the utterances are not only relevant enough to be worth the addressee’s attention, but they are also the most relevant ones compatible with the communicator’s abilities and preferences. To illustrate, consider the examples in (59)-(62).

- (59) To be collected. Spoke to waste services. (sign on a pile of rubbish outside a university department)
- (60) OK. Got that! (message on an ATM after the user has entered their PIN).
- (61) AB trying to find replacement cost for item returned damaged. Will contact when known. (Staff note on library user record. (AB = initials of a member of staff))
- (62) Very excited about entering our 2nd phase of 10 year plan.

In examples (59)-(62), it seems that resolving reference on a specific person or persons is not important. The exact identity of the agent of the verb does not contribute to the relevance of the utterance. Rather, the utterance achieves relevance by asserting that an action has taken (or will take) place. In (59), it is not necessary for the hearer to know or to be able to infer who has spoken to waste services. It is enough to know that the event has taken place. If the hearer follows the path of least effort, they will reach an interpretation which makes the utterance optimally relevant (i.e. that somebody has spoken to waste services, and hence that the pile of rubbish is being dealt with), and will therefore stop before they reach the point of assigning a precise reference to the null expression.²¹

In (60) and (61), the null subject is driven not by the unimportance of precise details about the identity of the referent, but by the speaker's preference for being non-committal. The ATM (or rather the human who programmed the ATM) in (60) is, we assume, aiming to give the transaction a friendly, relaxed and personal tone. However, using the personal pronoun 'I' in this utterance would anthropomorphize the ATM in a way that seems inappropriate when one is obviously dealing with a machine. The null subject version offers a compromise.²² Utterance (61) is a case of non-committal use in a different context. Although it may seem that the staff member 'AB' is the most accessible potential referent in this discourse context, (61) does not commit her to contacting the customer herself. Rather, the utterance goes no further than asserting that the user will be contacted. In this example, it is also unclear what the overt version of

²¹ We might achieve a similar effect by using the passive form, for example: Waste services have been informed.

²² I am not claiming that this was consciously thought out by the programmer of the machine, although this is not impossible.

the utterance would be (even if we assume co-reference with AB). We do not know if this was written by AB referring to herself in 3rd person, or by somebody else who is familiar with the situation. In terms of what is communicated, it does not matter. The utterance achieves relevance by asserting that the user will be contacted.

In utterance (62), it is again unclear what the overt version of the subject referring expression would have been. Assuming that a first person reading is highly accessible, a hearer still does not know whether the speaker is asserting (63) or (64). Again, it does not seem to matter in this case.

(63) We are very excited about entering our 2nd phase of 10 year plan.

(64) I am very excited about entering our 2nd phase of 10 year plan.

These examples highlight an important feature of the processing and interpretation of null subjects. Null subjects in English behave as referring expressions in their own right. It is not necessary to go through the process of identifying the content of the pronoun that has been elided. Rather, the hearer proceeds as with other referring expressions: he follows the path of least effort, looking for an overall interpretation that will satisfy his expectations of relevance. If identification of the agent is necessary (and as we have seen, this is not always the case), then he will test potential referents in order of accessibility. He need not identify a ‘missing’ pronoun, and reconstruct its encoded meaning before proceeding.

6.4.3 Null subjects and poetic effects

I am not presenting these three categories as necessarily an exhaustive taxonomy of English null subjects, or as theoretically distinct. They are simply illustrations of how a speaker can use null subjects as a form of referring expression, and thereby make her utterance optimally relevant. The cases may also overlap. So, for example, entries in the fictional *Bridget Jones’ Diary* by Helen Fielding were originally published as a serialised newspaper column, and as such were subject to strict word limits. As a result, whilst in some cases the nulls may contribute to the overall informal tone of the writing, they may also be one means of dealing with a length-restricted format, and their contribution to relevance may draw on both these aspects.

I have surveyed three broad categories of data, but my main claim is that null referring expressions are a means by which a speaker may make her utterance optimally

relevant. As a result, the null elements may interact with the discourse context in other ways, and achieve other effects. For example, the following passage is taken from the novel 'The Crow Road' by Ian Banks, and features several null subjects:

...I had to climb over the wall...gate was locked. By the time I got there he was out of reach. I thought he was shinning up a drainpipe. Just assumed. Heard rumbles, I think, but...didn't think anything of it. No flashes, that I can remember. Kenneth was yelling and swearing and shouting imprecations; calling down all sorts of punishments; I was trying to get him to come down; told him he'd fall; told him the police were coming; told him to think of his family. But he kept climbing.

(Banks, 1996, pp. 314-315).

Here the author uses the null subject utterances and short sentences to introduce a sense of panic and urgency into the narrative, perhaps indicating that the speaker is out of breath or under pressure. The fictional narrator is producing pressurized nulls, and so the author is able to create a feeling of tension and panic.

Similar effects are discussed by Blakemore (2008), who considers the effects produced by apposition of two or more identical or similar phrases in the representation of a character's thoughts, as in (65):

(65) In the beginning it was a tension, an element of strain that grew and crept like a thin worm through the harmony of their embrace (Hulme, 1985 [1984], p. 6) as quoted in (Blakemore, 2008, p. 37).

According to Blakemore, in this example, the repetition of the same idea, phrased in different ways, captures a difficulty in expressing a particular concept. However, she adds that crucially 'it is not Keri Hulme (the author) who is represented as grappling with her feelings...but the character whose feelings Hulme represents' (2008, p. 41). Likewise, it is not the author, Ian Banks, who is out of breath and panicking in the excerpt above, but the character, Prentice McHoan, as represented by Banks. Of course, these two examples differ in that Banks is writing a first person narrative, whereas Hulme is representing her character's thoughts in the free indirect style. However, Banks' first-person narrator is retelling past events, rather than relating them as they

happen, and the feelings and emotions represented are taken to be those of the character at the time of the events, not at the time of the description.

Thus, above and beyond the categories described in this section (pressurized, informal, ostensibly vague null subjects), null subjects may also be used in the creation of poetic effects. A writer/speaker may use them to create a particular style or tone by mimicking the sort of discourse contexts in which null subjects spontaneously occur, and thus creating a range of weak implicatures having to do with the character's emotions and state of mind (e.g. feeling pressurized, relaxed or evasive). In fictional texts, the reward for the extra effort expected of the hearer/reader may lie 'in the relationship which is created between the reader and the fictional individuals in the fictional world which the author is representing' (Blakemore, 2009).

6.5 Concluding remarks

In this chapter I have applied my account of referring expressions, developed in chapters 4 and 5, to the diary-style null subject data. I began by considering how null subjects in English might fit into other pragmatically-orientated analyses of referring expressions, and argued that neither Accessibility theory nor the Givenness hierarchy offers a satisfactory explanation of the data. To investigate null subjects further, I then considered how overtly realised pronouns function in discourse, and suggested a treatment of them within my relevance-based framework. Following earlier work by Wilson and Sperber (1993), Powell (2002) and Hedley (2005; 2007), I started from the assumption that pronouns encode procedures, and went on to consider what these procedures might look like. I argued that the gender, person and number information carried by pronouns should be seen as grammatical rather than semantic or conceptual, and that the procedures associated with pronouns work on a sub-personal level to indicate that the intended referent belongs to a particular category.

In the final section, I turned my attention back to the null subject data, and argued that the use of null subjects in certain discourse contexts in English creates extra or different effects which in turn justify the extra effort to which the hearer is put in processing null subject utterances. I identified three major categories of null subject use: informal style nulls, which contribute to a casual, intimate tone of a discourse, pressurized nulls, which free up restricted speaker resources, and so contribute to relevance, and ostensibly vague nulls, which save the hearer the effort of resolving reference when this does not directly contribute to the relevance of the utterance.

Finally, I noted that use of null subjects may be exploited by writers and speakers to create poetic effects.

Throughout this chapter, I have argued that the use of null subjects (and referring expressions in general) is driven by considerations of relevance. A speaker aims to produce an utterance which provides the most cognitive effects in her hearer, whilst minimising the effort to which that hearer is put, and allowing for her own abilities and preferences. In chapter 7, I apply this treatment to other, non-subject omissions, and to null subjects in other discourse contexts. I then consider the implications that my discussion may have for our understanding of referring expressions in general, and the conceptual/procedural distinction in particular.

Chapter 7: Conclusion

7.1 Introduction

In chapter 6, I returned to the diary drop null subject data in English and offered a relevance-based account of the phenomenon. I argued that the distribution and interpretation of null subjects falls out naturally in certain types of discourse context from the fact that the speaker is aiming at optimal relevance. In this concluding chapter I attempt to tie up some loose ends and draw some general conclusions.

In chapter 1, I discussed certain characteristics and distributional patterns associated with other null elements in the diary-style null subject examples, which include determiners, auxiliaries and prepositions, along with some objects. Following the analysis outlined above, these elements should only be omitted when they contribute little or nothing to the overall relevance of the utterance (in terms of either effort or effect). I do not pretend to be able to offer a comprehensive account of each of these categories in this thesis. However, in the next section I consider them briefly, and make some suggestions about how they might fit into my analysis of the pragmatics of diary-style texts. In the following section, I discuss the child null subject data and consider the status of null subjects in imperatives, which although fully grammatical in English, might be seen as ultimately motivated by similar considerations to those outlined above. I will also draw some general conclusions about the nature of the conceptual/procedural distinction, and end with some reflections on the differences between my approach to referring expressions in general, and null subjects in particular, and those proposed by Ariel and Gundel et al.

7.2 Non-subject omissions

7.2.1 Determiners

In chapter 4, I argued that the determiners in definite descriptions and complex demonstratives encode procedural meaning, and made some suggestions about how definite determiners might contribute to the overall interpretation of an utterance.

Recall that use of a plural noun phrase can lead to either a definite interpretation, as in (1), or a generic interpretation, as in (2), depending on whether or not a definite

determiner is present, while the combination of a plural noun with an indefinite determiner is ungrammatical.¹

- (1) James really likes the cats.
- (2) James really likes cats.
- (3) *James really likes a cats.

When the noun is singular, the contribution of the determiner to the interpretation of the referring expression is different.

- (4) James really likes the cat.
- (5) *James really likes cat.²
- (6) James really likes a cat.

As discussed in chapter 4, in interpreting ‘the cat’ in example (4), the hearer will test the most accessible potential referent from the set of things that are cats, and accept it if it leads to an overall interpretation that is relevant as expected. In this case, reference is resolved on one particular, definite cat. In (6), the indefinite determiner is generally seen as introducing a new discourse referent, and encouraging the hearer to open a new conceptual address or file in which subsequent information about this referent can be stored (Heim, 1982). As before, the conceptual information carried by the indefinite noun phrase will constrain potential discourse referents to the set of things that are cats. However, the hearer is not expected to identify the discourse referent with any particular, definite cat. In line with the definition of optimal relevance, there are three possible reasons why the speaker does not go further and direct the hearer to a more specific cat: either she does not know which cat it is that James likes (abilities), or she does not want to say which cat it is that James likes (preferences), or she simply does

¹ We can achieve an interpretation where James likes an undetermined sub-set of all cats by using the quantifier ‘some’:

- (i) James really likes some cats.

² There is, perhaps, an unusual interpretation available for (5) in which ‘cat’ behaves as a mass noun and has a meat-like interpretation. Compare with ‘James likes chickens’ and ‘James likes chicken’.

not regard it as relevant for the hearer to know which cat James likes. In the last case, the utterance will achieve relevance simply by informing the hearer that there exists a cat that James likes. In these circumstances, if the speaker were to formulate her utterance so that it leads the hearer to resolve reference on a particular cat, she would be putting him to extra effort, and encouraging him to look for extra effects to compensate.

Bearing this analysis in mind, let us return to the examples of null determiners in the diary-style texts, repeated here for convenience.

(7) Still have horrible cold.

(8) Very happy with dealer.

(9) Am delighted with dress.

(10) That would indeed be sick in manner of Oedipus.

(11) Also managed to conceal coat by rolling it into ball to create pleasing sense...

(12) Dialling tone seems normal.

Notice, first, that if the speaker wishes simply to avoid directing the hearer to a specific referent, she has the unmarked option of using the indefinite article, as in (13).

However, omitting the article altogether introduces a further element of vagueness or indeterminacy at the level of explicit content. In examples such as (7), it is not clear what exactly the overt form of the utterance would be if the determiner were to be articulated. For example, a fully articulated version of (7) could take any of the following forms.

(13) I still have a horrible cold.

(14) I still have my horrible cold.

(15) I still have that horrible cold.

The choice of one or other form does not increase the overall relevance of the utterance. By omitting the determiner, the speaker need not make this choice, and the hearer need not reconstruct the fully articulated sentence in order to retrieve the intended meaning. Thus, these types of determiner omission seem to function in a similar way to the ostensibly vague null subjects discussed above (section 6.4.2.3).

Null determiners can occur when the associated noun is intended to be understood as definite, as in (9), or indefinite, as in (11), and I suggest that they occur for similar reasons, and function in a similar way, to the null subjects. In some cases, the definite/indefinite information encoded by the determiner will make a crucial difference to the overall interpretation of the utterance. For example, when translating the phrase ‘the/a seagull’, in the play of the same name, from Russian to English, the choice between determiners makes a crucial difference to what the character Nina is understood to assert when she utters (16)/(17) (Frayn, 1986, pp. xix-xx).

(16) Nina: I am a seagull.

(17) Nina: I am the seagull.

Although both utterances are metaphorical, in (16) the metaphor attributes to Nina some properties of seagulls in general, whilst in (17) it attributes to her properties of a particular seagull that has featured earlier in the plot.

Whilst there are cases where the choice of determiner has a decisive effect on interpretation, in very many cases it will already be clear from the discourse context whether the speaker is referring to a particular definite category member or not. Thus, as with the third person gendered pronouns discussed above, there are occasions where the definite/indefinite information encoded by the determiner is largely redundant. For example, utterances (8) (‘Very happy with dealer’) and (9) (‘Am delighted with dress’) are from internet auction feedback message boards. It is clear from the discourse context that the speaker is giving feedback about a specific dealer and a specific dress respectively. In utterance (11) (‘Also managed to conceal coat by rolling it into ball to create pleasing sense...’), on the other hand, there is no identifiable ball in the discourse context to which the speaker might be referring, and the phrase achieves relevance merely by providing information about shape. However, in unmarked registers of English it is usual to mark definiteness on the determiner, just as it is usual to choose a gendered pronoun. Thus, following the same argument I used above for the pronominal

subjects, these redundant determiners will be high on the list of vulnerable items when it comes to saving time or space, or when seeking to create an informal tone or style.

7.2.2 Auxiliaries and the copula

The distribution of null versus overt forms of auxiliary verbs and the copula in English seems to further support the hypothesis that speech elements will be more vulnerable to omission the smaller their contribution to overall relevance. Occurring alongside the main verb in a sentence, auxiliaries in general provide information about the tense, aspect or voice of the verb, while modal auxiliaries carry information about the status of the embedded propositions as certainly, possibly or necessarily true. As noted in chapter 1, auxiliaries, in general appear vulnerable to omission in certain circumstances. My approach to diary-style omissions would predict that the smaller the contribution such elements make to the relevance of the utterance, the more likely it is that they may be omitted in these registers. This does indeed seem to be the case. For example, modal auxiliaries encoding epistemic or deontic information are not usually omitted.

(18) Get up in a minute.

(19) I must get up in a minute.

(20) I should get up in a minute.

(21) I will get up in a minute.

If a speaker were to utter (18), say, it would be very difficult for the hearer to infer which (if any) of the possibilities in (19)-(21) she was intending to communicate. Since the content of the modal makes a substantial difference to the relevance of the utterance, omission will be highly likely to lead to infelicity or misunderstanding. Moreover, omission of both the subject and the modal auxiliary yields the same form as the imperative, increasing the range of non-equivalent possible interpretations. The meaning encoded by the modal is therefore crucial to the overall relevance of the utterance, and it cannot be omitted. Utterances (22) and (23) are examples of overt modal auxiliaries from a diary-style text.

(22) But could buy book of cocktails (Fielding, 1996, p. 79).

(23) Must stop doing the instants (Fielding, 1996, p. 108).

For similar reasons, the auxiliary ‘to have’ is required to distinguish between the simple past tense and the perfect past tense. Its omission will change the overall interpretation of the utterance, and so it too will generally be overtly realised, as illustrated by (24).

(24) Have decided to serve the shepherds [*sic*] pie. (Fielding, 1996, p. 82)

Likewise, the so-called ‘dummy do’ auxiliary in constructions such as (25) is not usually dropped. This form not only carries tense information, but is only used in a positive declarative utterance when the speaker wishes to add emphasis. Since null elements are unable to carry stress, it is unclear how it could be dropped without changing the overall interpretation of the utterance.

(25) I did want to go.

(26) Did want to go.

(27) ?want to go.

(28) Wanted to go.

Although (27) is not ungrammatical in the diary register, it will lead to a present tense interpretation, and the variant in (28) does not carry the same emphasis.³

7.2.3 Objects

As outlined in chapter 1, diary-style utterances may contain instances of omitted objects as well as subjects, albeit much less frequently. My relevance-based account suggests an explanation both for the occurrence of null objects, and for the asymmetry in distribution. As is the case with subjects, objects will be vulnerable to omission in informal and pressurized registers when they contribute little or nothing to the overall

³ Notice also that although ‘I will’ and ‘I am’ are generally contracted to ‘I’ll’ and ‘I’m’ when spoken, they are vulnerable to subject-drop in diary-style written contexts (e.g. ‘Will call later’ ‘Am feeling under the weather today’). When written, the contracted form of ‘I am’ has one character more than the auxiliary on its own, but both forms of ‘I will’ are the same length. However, the contracted forms are part of the standard grammar of English and so are unmarked. My account therefore predicts that subject drop (as opposed to contraction) of these phrases will occur more often in informal style and ostensibly vague cases, where they lead to extra effects, than in pressurized contexts.

relevance of the utterance. Thus, in certain circumstances, and when the intended referent is the most accessible in the discourse context, the overt form of the object referring expression may be omitted. The asymmetry arises purely from the fact that subjects are more likely to be highly accessible than objects. According to my analysis, an object referring expression should only be vulnerable to omission if it refers to the most accessible potential referent in the discourse context. We would therefore normally expect to find an argument pronominalized when it is overtly realised in non-diary-style utterances. It follows that to study the rate of object drop in these registers, we should compare the frequency of null objects with the frequency of overt object pronouns. I assume that any non-pronominalised object referring expression encodes some other conceptual information that contributes to the overall relevance of the utterance. In a sample of 20 pages from a diary text featuring rampant subject drop, there were 3 overt object pronouns and 9 omitted object pronouns.⁴ This seems to suggest that the asymmetry is not the result of a grammatical restriction, but rather a matter of pragmatics. When an argument in object position would normally be realised as a pronoun, it will be omitted in 75% of cases in the informal and pressurised diary-style texts.⁵

As mentioned in chapter 1, it has been noted that recipes and other instructional registers often display a high rate of object drop in English (Massam & Roberge, 1989; Culy, 1996; Wharton, forthcoming), with utterances (29) and (30) being typical examples:

(29) Fry until golden brown.

(30) Drizzle with olive oil.

Although this type of procedural discourse is distinctive in style, I suggest that it is still pragmatically driven. For example, to be felicitous, the intended referents of the null

⁴ Sample taken from (Fielding, 1996). I did not include instances of reported speech.

⁵ I would suggest that the case for asymmetry is even weaker than this data suggests. The object pronouns which were overtly realised did not necessarily refer to a uniquely identifiable potential referent. For example:

(ii) If Magda's husband has nothing to be ashamed of dining with this worthless trollop in *my* suit, he will introduce me.

If the object pronoun 'me' were not overt, it would not be clear whether the speaker meant 'introduce me' or 'introduce her'. So the pronoun is overtly realised not because it is an object, but because it contributes to the overall relevance of the utterance.

objects in (29) and (30) must be highly accessible. Utterance (30) will not be felicitous if discourse initial, as there would be no accessible potential referent. Compare this with an instruction such as (31) on a bottle of medicine:

(31) Shake before use.

In (31) there is a highly accessible candidate referent (namely the bottle itself), but this is not the case in (30).

However, the recipe null objects do not obviously seem to fit with the categories of null subjects that I proposed in section 6.4.2. The discourse context is not particularly pressurized, the style is not informal, and the writers are certainly not intending to be vague about what should be fried, drizzled or shaken. How then might we account for these null objects? I suggest that in these cases the intended referents are the most accessible ones in the discourse context, and so the extra effort demanded of the hearer by leaving the object null is minimal. The nature of recipes (and instructional registers in general) means that the discourse follows a very specific and focused path within a fairly narrow discourse context, which could be in danger of becoming fairly repetitive. Compare (32) with (33):

(32) Take six onions. Chop them, salt them, drizzle them with oil, and fry them.

(33) Take six onions. Chop, salt, drizzle with oil and fry.

The use of overt pronominal objects in (32) does little if anything to reduce the hearer's effort (indeed, it could be argued that the intended referent, 'onions', is so highly accessible that the processing of the linguistic material puts the hearer to extra effort in (32)). Although repetition may be used by speakers to create extra or different effects (as discussed in Wilson & Sperber, 2004, pp. 219-222), this is not what the writer of (32) intends, and as no extra or different effects are obviously derivable, the utterance seems stylistically awkward⁶.

So the use of null objects in instructional registers such as recipes is once again a result of the speaker/writer aiming at optimal relevance.⁷

⁶ It may also be that the recipe null objects are a better candidate for the topic drop analysis than the diary-style null subjects.

⁷ Wharton (forthcoming) points out that recipes also include many null-subject utterances, and notes that apparently imperative sentences in recipes seem to lack directive force: their utterance is not 'an order, a

7.3 Further null-subject data

7.3.1 Embedded and root nulls

As noted by Haegeman and Ihsane (1999; 2001) and outlined in chapter 1, there appear to be two diary dialects in English. In one, null subjects are restricted to root positions, whilst in the other, embedded nulls can and do occur. Haegeman and Ihsane settle on a pronoun ellipsis account for the second, minority dialect and suggest that in such registers, ‘constraints on the identification of the null pronouns are relaxed’ (1999, p. 143). According to my relevance-based account, such differences should arise either because the restrictions on the abilities of the speaker are different, or because the different ‘dialects’ lead to extra or different effects. The stylistic variations between the two apparent dialects suggest that each of the motivating factors I have identified - familiarity, informality and pressure - forms a continuum, with different types of omission located at different points along the scale.

Haegeman and Ihsane first noted significant instances of embedded null subjects in informal, fictional diaries (for example Fielding, 1996; 2004). These texts are intended to simulate a private, intimate diary, and to create a sense of empathy and solidarity between the reader and the fictional narrator, Bridget Jones. The narrator is writing to and for herself, and the reader has the illusion of eavesdropping on this personal narrative. To create a realistic fictional diary in this style, the author, Helen Fielding, assumes almost complete mutuality between the cognitive environment of her narrator and that of the fictional addressee. Both are, after all, one and the same person. As a result, these diaries, along with other similar texts (Ginsburg, 1995; Bywater, 1998; Faulks, 1998) sit at the far end of the informality scale, and informal null subjects abound. This contrasts with non-fiction diaries written either explicitly for publication or at least with the possibility of a wider readership in mind.

It is not only the actual relationship between the hearer and speaker that drives the use and frequency of informal nulls (and other omissions), but the nature of the relationship as the author wishes to project it. The freedom with which the speaker can omit certain speech elements reflects how much she judges (or wishes her hearer to think she judges) to be shared between herself and her hearer. As noted in 6.4.3, this

request or an entreaty of any kind. It is simply one stage in a series of steps the reader can choose to follow or not’ (p.7). Also, some non-imperative recipe utterances may be subjectless, as in (iii) and (iv)

(iii) Serves six.

(iv) Feeds a family of four.

parallels the work on free indirect style by Blakemore (2009), in which the author aims to ‘present the illusion of a character acting out his mental state in an immediate relationship with the reader’. In the diary-style texts, the frequency of omissions will help to set the style and tone of the discourse by giving access to a wide range of weak implicatures.

However, informality is not the only motivating factor which involves a continuum. Discourse contexts may also be more or less pressurized, and this in turn will affect the likelihood and frequency of omissions. The omission of prepositions (although attested) is much less common than the omission of the other parts of speech discussed above. However, prepositions are frequently omitted in telegrams. Telegram companies charged for their service by the word, and so they have unwittingly provided us with examples of utterances produced in an explicitly and artificially pressurised discourse context. Whilst the information encoded by prepositions may contribute sufficiently to the relevance of an utterance to be articulated in most of the diary-style texts, they become vulnerable once the pressure on word limit is increased to this degree. In each case, the utterance produced is still the most relevant one that the speaker could have uttered given her abilities and preferences. However, an increase in the pressure on the speaker, affecting her abilities and/or preferences, means that the hearer might expect to have to do more inferential work, and thus expend more effort to derive the intended meaning.

7.3.2 Child nulls again

In chapter 2, I outlined my argument that null subjects in child language result from the fact that the child is operating under processing constraints, and therefore omits the elements which contribute least to the overall relevance of an utterance. In chapter 6, I made a parallel case for the adult nulls in diary-style texts, and discussed why it is that pronouns often contribute the least in relevance-terms.

Whereas in the adult cases I have identified three circumstances which might lead to the use of null subjects, in the case of the child null subjects, the child’s developing, and thus limited, processing and cognitive capabilities, are the main driving force. As Lois Bloom puts it, the truncated child utterances ‘can be compared to the kinds of sentences produced by an adult who is under pressure to be brief’ (Bloom, 1970, p. 139). Whereas the pressurized nulls in the adult examples may result from time or space constraints, the child going through the null subject stage of acquisition is subject

to the internal pressures and restrictions of her developing cognitive, processing and linguistic systems.

A relevance-based account has the advantage of being able to explain many of the distributional properties and other characteristics associated with the child nulls, as outlined in chapter 2. As discussed there, a major objection to a performance-driven analysis of child null subjects is the fact that such approaches are allegedly unable to account for the observed object/subject asymmetry. Competence accounts (Hyams, 1986; Hyams & Wexler, 1993; Rizzi, 1994; 2005; Bromberg & Wexler, 1995) can specify that omissions are restricted to subject position, but this option is not available to performance accounts. However, the asymmetry is to be expected on a relevance-based account. As with adult null subjects, there are likely to be more occasions when the subject referent is more accessible than the object referent, and will therefore be the first to be tested, than occasions when the converse is true.

The correlation between the end of the null subject stage and the acquisition of inflection is also predicted on my account. When processing resources are limited, we would expect not only subjects but other high-cost, low-effect elements to be vulnerable. The lack of grammatical inflection is less likely to interfere with the overall interpretation of an utterance than, say, the lack of a content word such as a noun or verb⁸. The correlation between expletive use and the end of the null subject phase is likewise expected on my relevance-based analysis. If the elements which contribute least to the overall relevance of an utterance are most likely to be omitted, then expletive pronouns should be highly vulnerable. On my account, then, expletives are not the catalyst for grammatical change, as proposed by Hyams (1986). Rather, once the child's processing capacities have matured to the point where expletives are no longer dropped, we would expect to find that other omissions are also no longer necessary. As discussed above, Valian (1990) finds evidence of expletive use across the MLU range of children she studied, both within and after the null subject phase. We can understand why this might be so if we treat the omissions as due to a processing overload (and therefore as linked to the child's abilities) rather than as related to the acquisition of expletives as a grammatical category. It is not that the child cannot represent or produce an expletive during this phase. Rather, expletives are highly vulnerable to omission when the child's processing capacity is limited.

⁸ I am not claiming that the child has already acquired inflection and then omits it. A model where inflection is mastered late because the child pays less attention to this aspect of utterances than to others is also compatible with my analysis.

The empirical findings discussed in Valian, Aubry and Hoeffner (1996) are also consistent with my relevance-based analysis. They found that in imitation experiments, children in the null subject stage imitated pronominals less often than full subjects. If a referent is pronominalized in the target sentence, then we can assume that it is highly accessible in the discourse context. It is therefore much more likely to be vulnerable to omission than, say, a definite description which encodes conceptual information, and therefore contributes to reference resolution. That expletive pronouns will be omitted more often than referential pronouns is to be expected for the same reasons.

Relevance theory thus gives us a means to reinterpret the intuitive, but at times vague, conclusions drawn by the processing and pragmatic accounts within an independently motivated cognitive pragmatic framework. Relevance theory explains why the sort of elements which Bloom describes as ‘weakly stressed, carry little information value, [and] are most predictable’ (1970, p. 140) are those which are most likely to be omitted.

7.3.3 Imperatives

There is, of course, a category of null subject sentences in English which are fully grammatical across all registers and dialects: imperatives. Traditionally, the subjects of imperative sentences have been analysed as second person (Bullokar, 1586; Jespersen, 1940; Sweet, 1960; Katz & Postal, 1964; Thorne, 1966; Schmerling, 1982).

(34) Get up.

(35) You will get up.

(36) Get up, will you?

On these analyses, the subject and modal in (35) are unarticulated in (34), and they are moved to produce the tag question in (36). The short forms are seen as acceptable because ‘the subject *you* is “understood”, not needing to be marked because it is the only possible subject’ (Thorne, 1966, p. 69).

However, more recent approaches to imperatives have noted that a range of other subjects are available (Platzack & Rosengren, 1998; Jensen, 2003).

(37) Nobody move.

(38) Everybody stay where you are.

(39) Whoever laughed leave the room now.

(40) Somebody help me.

Although it is outside the scope of this thesis to offer a syntactic analysis of the English imperative, I feel that the parallels between the subjectless imperative sentences and the diary-style subject deletions are worthy of more comment. Platzack & Rosengren (1998) refer to the noun phrase which is optionally overtly realised in imperatives as overt ImpNP, and go on to claim that ‘as in the case of null subjects in finite clauses, the overt realization of ImpNP is a matter of pragmatic considerations’ (1998, p. 197). I want to suggest that these pragmatic considerations, as with the diary nulls, are ultimately considerations of relevance in Sperber and Wilson’s sense.

Since imperatives are typically used to convey requests, orders or instructions, they are generally addressed directly and explicitly to a hearer or group of hearers. Thus, the mere act of addressing someone with an imperative utterance will, in general create the expectation that they are the subject of the request, order or instruction.⁹ It is highly likely that the addressee will also be indicated by some extra-linguistic means (eye direction, head nodding, pointing or other gesture). If the speaker chooses to overtly articulate the subject, she is therefore, in effect, requiring the hearer to expend more effort than is necessary, and so the hearer will be entitled to expect extra effects. Thus, the bare imperatives are the unmarked versions in most cases.

Given this approach to the pragmatics of imperatives, it is easy to see why some existing accounts claim that the subject of an imperative sentence must be in the second person. This will indeed be the most frequent interpretation, and evidence from tag questions suggests that even examples such as (37)-(39) are interpreted as second person, as illustrated in (41) and (42):

(41) Everybody stay where you are, will you?

(42) Whoever laughed leave the room now, will you?

⁹ Mauck, Pak, Porter and Zanuttini (2005, p. 146) present a cross-linguistic study of imperative subjects, and conclude that whilst the subject and addressee need not always coincide, at least ‘In all cases, the addressee must have some control over whether the subject has the property denoted by the predicate’.

Furthermore, these examples lend support to the hypothesis that the choice between overt and null imperative subjects is pragmatically driven. It seems that null subjects are not always a felicitous option for imperative utterances. Consider (37) ('Nobody move'). Uttering this sentence without the overt subject 'nobody', as in (43), will lead to a very different interpretation.

(43) Move.

It is not possible to address an utterance or other ostensive stimulus to nobody. It might be argued that this is because not only the subject but also the sentence negation have been omitted, and that the grammatically correct null subject version of (37) would therefore be as in (44), with the missing subject understood as 'everybody'.

(44) Don't move.

But here again, (44) has a range of possible interpretations on which the missing subject would be a single person or a few people rather than everyone in the group of addressees. Thus the act of ostensive communication alone is not enough to guide the hearer(s) to the intended referent, and to achieve optimal relevance, the speaker must articulate the subject. The extent to which the overt subjects in (38) and (39) may be left unarticulated depends on the discourse context in which they are uttered. As always, the hearer will test interpretations in order of accessibility. If the intended referents of 'everybody' and 'whoever laughed' are the most accessible potential referents in their respective discourse contexts, then the option to utter a bare imperative may be available to the speaker. And, as always, accessibility may be affected by, amongst other factors, the prior discourse, the physical context and any extra-linguistic gestures of the speaker.

Utterance (40) ('Someone help me'), and its subjectless alternative in (45), again shed some light on the role of pragmatics in the production and interpretation of imperatives.

(45) Help me.

Whilst a second person interpretation may be the most common for imperative utterances, examples such as (45) suggest that in some cases, as with the ostensibly

vague diary null subjects, it might not be necessary to reconstruct a determinate subject referent at all. The relevance of an utterance of (45) is most likely to arise from the fact that the speaker is requesting help. It is not necessary for a hearer to decide whether the speaker intended to produce an abbreviated version of (40), or any of a range of other possible alternatives such as (46)-(48).

(46) You help me.

(47) John help me.

(48) One of you help me.

The intended addressee (and subject) of the utterance may be obvious from the discourse context or extra-linguistic cues, but regardless of this, the hearer is likely to have achieved adequate effects before he goes to the effort of resolving the reference of the missing subject. Alternatively, the appeal to the speaker's abilities and preferences in the presumption of optimal relevance might also play a part in the processing of (45). The speaker might not have a preference about who helps her, as long as someone does.

Bearing this analysis in mind, consider the following instructional imperatives:

(49) Stand behind the yellow line.

(50) Mind the gap.

(51) Shake before use.

(52) Love like you've never been hurt.

Again, with examples (49)-(52), it is not necessary to reconstruct a subject (second-person or other) in order to interpret the intended meaning. In a similar fashion to the ATM example discussed in section 6.4.2.3 ('OK, got that'), the speakers of these utterances are being deliberately vague about what the subject of these utterances is intended to be. It is not clear how such examples would be enriched to include overt subjects. Examples (53)-(56) suggest some possibilities for utterance (49).

(53) You should stand behind the yellow line.

(54) Everyone should stand behind the yellow line.

(55) Passengers should stand behind the yellow line.

(56) Anyone standing on this platform should stand behind the yellow line.

None of these, however, can be said to exactly capture what is communicated by (49). Rather, it might be said that the utterance is addressed to whoever might read it and find it relevant enough to be worth their attention. It would be extremely hard to spell out more fully precisely who this is: for instance, it is clearly addressed not only to passengers but to people accompanying or meeting passengers, people walking along the platform to another destination or people who have wandered into the station by mistake, but not to passengers getting on or off the train, people repairing the track, people rescuing children who have toddled over the line, etc. etc. Thus, a huge amount of contextual information is needed to assign a referent appropriately in a null subject imperative such as this.

Finally, there is the issue of subjectless tag questions such as (57)-(58)

(57) Sit down, will you?

(58) Feeling sick, is he?

Given my approach to subjectless utterances, and imperatives in particular, it seems that the addition of a tag question may add emphasis, as in (57), or may act as a repair mechanism to avert possible misunderstanding, as in (58).

7.4 Concluding remarks

7.4.1 Revisiting procedural meaning

7.4.1.1 The vulnerability of procedural meaning to omission

In the course of this thesis, I have argued that the omission of subjects and other material in diary-style texts can be explained by taking a relevance-based pragmatic approach. I have proposed a procedural analysis of pronouns, and linked their

procedural nature to the fact that they are vulnerable to omission in the circumstances I have identified, and for the reasons I have discussed. In each case, the omissions result from pressure on the speaker's abilities or preferences, or provide a means for achieving extra or different effects.

An interesting implication of my analysis is that expressions which encode mainly procedural meaning should be particularly vulnerable to omission. This follows naturally from the nature of procedural meaning and its role within the relevance-theoretic pragmatic framework. As Carston (2002b) puts it, the role of procedural meaning is to 'indicate, guide, constrain, or direct the inferential phase of comprehension', which makes it essentially an 'effort-saving device' (2002b, p. 162). Procedural meaning contributes to the overall relevance of an utterance by reducing hearer's effort. Bearing this in mind, let us return to the three categories of null subjects discussed above.

In pressurised environments (e.g. newspaper headlines, message boards, postcards, text messages, note-taking etc) the speaker's production abilities are constrained. In these discourse contexts, it will be manifest to the hearer that the speaker is operating under constraints, and that he should adjust his expectations of relevance accordingly. Whilst he can still presume that the speaker has aimed at optimal relevance within the constraints imposed by her limited production abilities, the level of relevance achieved may be relatively lower than if the utterance were produced without constraints. In other words, he must expect to put in more effort in order to derive the same overall interpretation, and this demand for extra effort, rather than contributing to extra effects, must simply be forgiven or condoned. Omission of the procedurally encoded elements is the best strategy for the speaker. It might put the hearer to more effort, but the message can still be successfully conveyed, and should still satisfy the hearer's (reduced) expectations of relevance. This approach ties in with a more general observation that a hearer's expectation of relevance will vary from speaker to speaker and from circumstance to circumstance. So, for example, we are unlikely to expect the same level of effect, or the same balance of effort and effect, when speaking to a child as we do when speaking to a university professor. Thus, the 'abilities and preferences' clause of the presumption of optimal relevance applies to both effort and effects. A hearer is only entitled to presume that an utterance will provide the most effects compatible with the speaker's abilities and preferences (i.e. what she knows, and is prepared to divulge), and that it will put him to the minimal

effort compatible with the speaker's abilities and preferences (i.e. that it is the most economical one that she is willing and able to provide in the circumstances).

In the informal diary-style discourse contexts, there may be no obvious pressure on the speaker's time or restriction on the length of utterance, and yet speakers still omit potentially effort-saving linguistic devices. Again, the hearer will assume that the speaker is being optimally relevant, and therefore that she did not intend to put him to gratuitous effort. In this case, she had the ability to use the effort-saving devices, but preferred not to. As a result, the hearer must assume that she intended the extra effort required to yield some extra or different effects. This assumption in turn suggests that the speaker trusted the hearer to derive the intended interpretation without further linguistic cues, resulting in the informal and intimate tone of the diary-style nulls.

Finally, what role does the (absence of) procedural meaning play in the case of the ostensibly vague nulls? I suggest that it plays no role. The hearer will follow the path of least effort in deriving possible interpretations, and will stop when her expectations of relevance are satisfied. With the ostensibly vague nulls, the hearer's expectations of relevance will be satisfied before he has reached the point of resolving reference in any but the most general terms. In these circumstances, a procedural device which 'indicates particular computational processes' would not be effort saving, but would instead put the hearer to gratuitous effort. Thus each of the uses of null subjects that I have identified ultimately arises from the speaker's aim of achieving optimal relevance.

7.4.1.2 The nature of procedural meaning

My analysis of null subjects in English, and referring expressions in general, also raises interesting issues about the nature of procedural meaning and its relation to conceptual meaning.

In chapter 6, I discussed the gender, number and person information encoded by pronouns, and argued that it should be seen as purely grammatical, rather than as properly semantic or conceptual. I also linked this suggestion to the more general claim that procedural information functions on a sub-personal rather than personal level. On this account, the procedures encoded by pronouns and demonstratives do not restrict the set of potential referents by imposing conceptual constraints (e.g. MALE, INANIMATE, PROXIMAL) as suggested by Zaki (2009) and Hussein (2008), but by adding an extra layer of activation to a category of items identifiable only in sub-personal terms. For instance, the pronoun 'she' might be seen as adding an extra layer of activation to the category of items that can also be legitimately referred to using the pronoun 'she',

regardless of whether they fall under the concepts ANIMATE or FEMALE; the pronoun ‘that’ might be seen as activating a complex computation involving assessments of the relative positions of items to a deictic centre along a variety of dimensions, and so on.

We might see some parallels between this proposed move away from semantic or conceptual treatments of pronouns and determiners and recent moves away from attempts to define grammatical categories in semantic or conceptual terms. For instance, defining a noun as an expression which denotes a person, place or thing is generally regarded as falling far short of reflecting a native speaker’s grammatical knowledge. Rather, nouns are typically characterised in functional terms, as words which a native speaker can use in a certain way, which provide input to certain types of syntactic computation, and are therefore in some sense defined by the native speaker’s tacit grammatical knowledge. Thus, although the set of nouns will generally include items denoting people, places and things, it cannot be defined as such. Similarly, although the set of referents the pronoun ‘she’ can be used to pick out will generally include items that fall under the concepts ANIMATE and FEMALE, it cannot be defined as such. A similar argument can be made for the proximity information encoded by demonstratives.

In moving towards a sub-personal account of the procedural information encoded by pronouns and other referring expressions, I am building on a proposal by Wilson (2009), that procedural information in general should be analysed in sub-personal rather than personal terms. Developing ideas put forward in Sperber & Wilson (1986/1995, pp. 172-3), she suggests that the well-formed formulas in a language may get their meanings in three distinct ways: by being translated into expressions of another language, by corresponding systematically with states of the language user, or by corresponding systematically with states of affairs in the world. In her view, conceptual expressions get their meanings in the first of these ways, by being translated into the language of thought, procedural expressions get their meanings in the second of these ways, by systematically activating states of the language user, and expressions in the language of thought get their meanings in the third of these ways, by corresponding systematically with states of affairs in the world. She concludes, ‘We may expect procedural items (e.g. pronouns, particles, interjections) to activate procedures formulated in a sub-personal “machine language” rather than full-fledged concepts which are constituents of a “language of thought” and thus available for general inference (cf. pronouns).’ In developing my procedural account of pronouns and referring expressions, I have tried to flesh out these ideas to some extent.

More generally, a sub-personal approach to procedural meaning might shed new light on the comment made at several points in this thesis, and in the literature on the conceptual-procedural distinction in general, that while conceptual meaning is relatively easy to paraphrase in more or less adequate intuitive terms, procedural expressions such as ‘but’, ‘so’, ‘after all’, ‘still’, etc. present a much greater challenge. This is generally explained by appealing to the distinction between representation and computation, and noting that while conceptual expressions encode concepts that figure directly in the language of thought, procedural expressions encode computations, which are not so easily brought to consciousness. However, as the literature on pronouns shows, it is possible to combine a procedural approach to pronouns with the claim that the computations involved pick out sets of referents that fall under regular concepts such as MALE, or ANIMATE, made available by the language of thought. However, if we take seriously the claim that procedural meaning is sub-personal rather than personal, this approach becomes harder to justify. I hope to pursue this aspect of my analysis in future work.

7.4.2 In support of a relevance-based account of referring expressions

There are several key differences between my account of referring expressions, and those provided by the other cognitive approaches I have discussed (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993; Gundel & Mulkern, 1998), not only in the analysis of null subjects in English, but in the treatment of referring expressions in general.

In section 7.2.2, I touched on the role of ‘dummy-do’ auxiliaries, and noted that they cannot generally be dropped without affecting the interpretation of an utterance. These data highlight an interesting general distinction between cases where there is a less costly grammatical alternative, and those in which there is not. The dummy auxiliary ‘do’ only adds emphasis in positive declarative utterances such as (25) (‘I did want to go’), where there is an optional alternative without stress, in this case (59).

(59) I want to go.

The ‘do’ auxiliary does not carry stress in negatives, as in (60), or interrogative, as in (61).

(60) I do not want to go

(61) Do you want to go?

A relevance-theoretic account allows us to distinguish between these cases. When there is a less costly grammatical alternative, as with (25), extra effort demanded implies extra or different effects, but when there is no less costly grammatical alternative, as with (60) and (61), however much effort is demanded, the hearer is not entitled to look for extra effects. This same general point applies to the optional overt subjects in null subject languages, where there is always a less costly grammatical alternative to a sentence with an overt pronominal. As a result, when an overt pronoun is used in these languages, the hearer is entitled to look for extra effects (see footnote 19, section 6.4.2.1). It is hard to see how a strictly grammatical account of null subjects could deal with such cases.

As for referring expressions in general, on my account they are not seen as encoding semantic information about the Accessibility or cognitive status of the intended referent, and they do not instruct the hearer where in the search space he should look. Instead, I have taken seriously the idea that a hearer following the relevance theoretic comprehension procedure will test possible interpretations in order of accessibility. I have therefore argued that referring expressions should be seen as encoding information that restricts the set of potential referents to a point where the intended referent is, if all goes well, the most accessible one that yields an interpretation that is relevant in the expected way.

Rather than defining a referring expression by its form (pronoun, proximal complex demonstrative, long definite description etc) and locating it on a scale or hierarchy, I have argued that each referring expression may encode conceptual and/or procedural meaning. This meaning may contribute to reference resolution and hence to the explicit content of an utterance, but it may also affect what is implicitly communicated by an utterance, by contributing to the derivation of implicatures and other inferential effects. Thus, choice of referring expression is seen as playing a role in all three of the sub tasks which, according to relevance theory, contribute to the overall interpretation of an utterance.

Crucial to my account is the role of the discourse context. In previous cognitive-pragmatic approaches to referring expressions, the importance of the context as a source both of candidate referents and of potential premises for inference is overlooked or remains fairly peripheral to the analyses. On my account, referring expressions are seen as functioning relative to the discourse context. Moreover, context selection, guided by

the presumption that the utterance will be optimally relevant, plays a role not only in securing reference but also in deriving contextual implications and other cognitive effects.

A major advantage of my approach to referring expressions is that it allows for a unified account, in which Ariel's 'special uses' and Lakoff's 'emotional' demonstratives are no longer seen as exceptional or unexpected. Rather, they fall out naturally from the interaction between encoded information, the discourse context and the speaker goal of achieving optimal relevance.

7.4.3 Final remarks

In this thesis, I set out to provide a relevance-based account of the pragmatics of diary-style null subjects in English. As discussed in chapter 1, although a variety of grammatical analyses have been proposed for these data, all acknowledge that pragmatic factors play a decisive role in their use and distribution. I have attempted to offer a coherent account of what these pragmatic factors are, based on an independently motivated pragmatic framework, relevance theory. In doing so, I have outlined my relevance-based approach to the use and interpretation of referring expressions in general. On my account, null subjects in English are not as exceptional as they may at first appear. Rather, they are just one means by which a speaker may achieve her overall aim of producing an optimally relevant utterance: that is, an utterance which is at least relevant enough to be worth the hearer's attention, and which, moreover, produces the greatest effects, for smallest effort, compatible with her own abilities and preferences. As such, they should be expected to arise spontaneously in the types of discourse context I have described, and for the reasons I have attempted to outline.

References

- Adger, D., & Harbour, D. (2008). Why Phi? In D. Harbour, D. Adger, & S. Bejar (Eds.), *Phi Theory: Phi-Features Across Modules and Interfaces* (p. 1034). Oxford: Oxford University Press.
- Agar, N. (1993). What do frogs really believe? *Australasian Journal of Philosophy* , 71 (1), 1-12.
- Allen, S. E. (2000). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics* , 38 (3), 483-521.
- Anderson, S. R., & Keenan, E. L. (1985). Deixis. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (Vol. 3, pp. 259-308). Cambridge: Cambridge University Press.
- Ariel, M. (1985). Givenness Marking. *PhD Thesis* . Tel-Aviv University.
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics* , 24 (1), 65-87.
- Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. London: Routledge.
- Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics* , 16, 443-463.
- Ariel, M. (1994). Interpreting anaphoric expressions: a cognitive versus a pragmatic approach. *Journal of Linguistics* , 30 (1), 3-42.
- Ariel, M. (2001). Accessibility theory: an overview. In T. Sanders, J. Schliperoord, & W. Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (pp. 29-87). Amsterdam: John Benjamins.
- Bach, K. (1992). Intentions and demonstrations. *Analysis* , 52, 40-146.
- Banks, I. (1996). *The Crow Road*. London: Abacus.
- Black, C. A. (1996). A step-by-step introduction to Government and Binding theory of syntax. *Notes on Linguistics* , 73, 5-12.
- Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.
- Blakemore, D. (1992). *Understanding Utterances*. Oxford: Blackwell.
- Blakemore, D. (1993). The relevance of reformulations. *Language and Literature* , 2 (2), 101-120.
- Blakemore, D. (1994). Relevance, poetic effects and social goals: a reply to culpeper. *Language and Literature* , 3 (1), 49-59.
- Blakemore, D. (2000). Indicators and procedures: nevertheless and but. *Journal of Linguistics* , 36, 463-486.
- Blakemore, D. (2002). *Relevance and Linguistic Meaning: the Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.

- Blakemore, D. (2007). Constraints, concepts and procedural encoding. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 45-66). Basingstoke: Palgrave MacMillan.
- Blakemore, D. (2008). Apposition and affective communication. *Language and Literature* , 17 (1), 37-58.
- Blakemore, D. (2009). The use of audience-directed expressions in indirect thought representations. *Procedural Meaning: Problems and Perspectives*. Madrid.
- Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. Cambridge, MA: MIT Press.
- Bloom, L., Lightbrown, P., & Hood, L. (1975). Structure and Variation in Child Language. *Monographs of the Society for Research in Child Development* , 40 (2).
- Bloom, P. (1990). Subjectless sentences in Child Language. *Linguistic Inquiry* , 21 (4), 491-504.
- Bloom, P. (1993). Grammatical continuity in language development: the case of subjectless sentences. *Linguistic Inquiry* , 24 (4), 721-734.
- Bromberg, H. S., & Wexler, K. (1995). Null subjects in child WH-questions. (C. T. Shutze, J. B. Ganger, & K. Broihier, Eds.) *Papers in Language Processing and Acquisition* , 26, pp. 221-247.
- Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review* , 34, 133-151.
- Brown, R., & Fraser, C. (1963). The acquisition of syntax. In C. N. Cofer, & B. S. Musgrave (Eds.), *Verbal Behaviour and Learning*. New York: McGraw-Hill.
- Bullokar, W. (1586). *Pamphlet for Grammar*. London: Edmund Bollifant.
- Bywater, M. (1998, June 7). Mark Darcy's Diary. *The Observer* .
- Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behaviour* , 16, 601-609.
- Carston, R. (1997). Informativeness, relevance and scalar implicature. In R. Carston, & S. Uchida (Eds.), *Relevance Theory: Applications and Implications* (pp. 179-236). Amsterdam: John Benjamins.
- Carston, R. (2002a). *Thoughts and Utterances : the Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Carston, R. (2002b). Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind & Language* , 17, 127-148.
- Carston, R. (2004). Relevance theory and the saying/ implicating distinction. In L. Horn, & G. Ward (Eds.), *Handbook of Pragmatics*. Oxford: Blackwell.

- Carston, R. (2005). A Note on Pragmatic Principles of Least Effort. *UCL Working Papers in Linguistics* , 17, 271-278.
- Cheshire, J. (1996). That jacksprat: an international perspective on English that. *Journal of Pragmatics* , 25, 369-93.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge MA: MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris: Dordrecht.
- Chomsky, N. (1982). Some Concepts and Consequences of the Theory of Government and Binding. Cambridge, Mass.: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge MA: MIT Press.
- Clahsen, H. (1986). Verb inflections in German child language: Acquisition of agreement markings and the functions they encode. *Linguistics* , 24, 79-121.
- Clancy, P. M. (1993). Preferred argument structure in Korean acquisition. In E. V. Clark (Ed.), *Proceedings of the 25th Annual Child Language Research Forum* (pp. 307-314). Stanford, CA: CSLI.
- Clancy, P. M. (1997). Discourse motivations of referential choice in Korean acquisition. In H. Sohn, & J. Haig (Eds.), *Japanese / Korean Linguistics 6* (pp. 639-659). Stanford, CA: CSLI.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi (Ed.), *Elements of Discourse Understanding* (pp. 10-63). Cambridge: Cambridge University Press.
- Corazza, E., Fish, W., & Gorvett, J. (2002). Who is I. *Philosophical Studies* , 107, 1-21.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Cornish, F. (2001). "Modal" that as determiner and pronoun: the primacy of the cognitive-interactive dimension. *English Language and Linguistics* , 5 (2), 297-315.
- Culy, C. (1996). Null objects in English recipes. *Language Variation and Change* , 8, 91-124.
- Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge and Kegan Paul.
- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Dowty, D., & Jacobson, P. (1988). Agreement as a semantic phenomenon. In J. Powers, & F. de Jong (Eds.), *Proceedings of the Fifth ESCOL* (pp. 95-101). Ithaca: Cornell Linguistics Club.
- Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology* , 32, 247-255.

- Elton, M. (2000). The personal/sub-personal distinction: An introduction. *Philosophical Explorations* , 3 (1), 2-5.
- Evans, G. (1980). Pronouns. *Linguistic Inquiry* , 11 (2), 337-362.
- Faulks, S. (1998, Jan. 7). Jack the Lad. *The Guardian* .
- Fielding, H. (1996). *Bridget Jones' Diary*. Oxford: Picador.
- Fielding, H. (2004). *Bridget Jones: the Edge of Reason*. Oxford: Picador.
- Fillmore, C. J. (1982). Towards a descriptive framework for spatial deixis. In R. J. Jarvella, & W. Klein (Eds.), *Speech, Place and Action: Studies in Deixis and Related Topics* (pp. 31-59). Chichester: Wiley.
- Fillmore, C. J. (1997). *Lectures on Deixis*. Stanford, CA.: CSLI Publications.
- Fodor, J. (1975). *The Language of Thought*. Cambridge, Mass.: Harvard University Press.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, Mass: MIT Press.
- Fodor, J. (2008). *LOT2: The Language of Thought Revisited*. Oxford: OUP.
- Frayn, M. (1986). Introduction to *The Seagull* by Anton Chekhov. London: Methuen.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry* , 5, 459-464.
- Garvey, C., Caramazza, A., & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. *Cognition* , 3, 227-243.
- Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language* , 30, 431-451.
- Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language* , 27, 699-717.
- Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge University Press.
- Ginsburg, A. (1995). *Journals 1954-1958*. (G. Ball, Ed.) London: Penguin.
- Glover, K. D. (2000). Proximal and distal deixis in negotiation talk. *Journal of Pragmatics* , 26, 915-926.
- Goldin-Meadows, S., & Mylander, C. (1984). Gestural communication in deaf children: the effects and non effects of parental input on early language development. *Monographs of the Society for Research in Child Language Development* (49).
- Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development*. London: Academic Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and Semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). *Centering: A Framework for Modelling the Local Coherence of Discourse*. Institute for Research in Cognitive Science. Philadelphia: University of Pennsylvania.
- Guilfoyle, E. (1984). The Acquisition of Tense and the Emergence of Lexical Subjects in Child Grammars of English. *Boston University Conference on Language Development*. Boston.
- Guilfoyle, E., & Noonan, M. (1989). Functional Categories and Language Acquisition. *Unpublished manuscript*. McGill University.
- Gundel, J. K. (2003). Information structure and referential givenness/newness: How much belongs in the grammar? In S. Muller (Ed.), *Proceedings of the HPSG03 Conference*.
- Gundel, J. K., & Mulken, A. E. (1998). Quantity implicatures in reference understanding. *Pragmatics and Cognition*, 6 (1/2), 21-45.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69 (2), 274-307.
- Haan, G. J., & Tuijnman, K. (1988). Missing subjects and objects in child grammar. In P. Jordens, & J. Lalleman (Eds.), *Language Development* (pp. 101-122). Dordrecht: Foris.
- Haegeman, L. (1990a). Understood subjects in English diaries: on the relevance of theoretic syntax for the study of register variation. *Multilingua*, 9 (2), 157-199.
- Haegeman, L. (1990b). Non-overt subjects in diary contexts. In J. Mascaro, & M. Nespore (Eds.), *Grammar in Progress: GLOW Essays for Henk van Reimsdijk* (pp. 167-179). Dordrecht: Foris.
- Haegeman, L. (1994). *Introduction to Government and Binding Theory* (Second edition ed.). Oxford: Blackwells.
- Haegeman, L. (1997). Register variation, truncation and subject omission in English and French. *English Language and Linguistics*, 1, 233-270.
- Haegeman, L. (2000). Adult null subjects in non-pro-drop languages. In M. Friedemann, & L. Rizzi (Eds.), *The Acquisition of Syntax: Studies in Comparative Development Linguistics* (pp. 129-169). Harlow: Pearson.
- Haegeman, L., & Gueron, J. (1999). *English Grammar: a Generative Perspective*. Oxford: Blackwell.

- Haegeman, L., & Ihsane, T. (1999). Subject ellipsis in embedded clauses in English. *English Language and Linguistics* , 3 (1), 117-145.
- Haegeman, L., & Ihsane, T. (2001). Adult null subjects in the non-pro-drop languages: two diary dialects. *Language Acquisition* , 9 (4), 329-346.
- Hall, A. (2004). The meaning of but: A procedural reanalysis. *UCL Working Papers in Linguistics* , 16, 199-236.
- Hall, A. (2007). Do discourse markers encode concepts or procedures? *Lingua* , 111 (1), 149-174.
- Halliday, M. A. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold Ltd.
- Hamann, C., & Plunkett, K. (1997). Subject omission in child Danish. Proceedings of the Annual Boston University Conference on Language Development , 21, 220-232.
- Hamann, C., & Plunkett, K. (1998). Subjectless sentences in child Danish. *Cognition* , 69, 35-72.
- Hamann, C., Rizzi, L., & Frauenfelder, U. (1996). On the acquisition of subject and object clitics. In H. Clahsen, *Generative perspectives on language acquisition*. Amsterdam: John Benjamins.
- Hedley, P. (2005). Pronouns, procedures and relevance theory. *Durham Working Papers in Linguistics* , 11, pp. 41-55.
- Hedley, P. (2007). Anaphora, Relevance and the Conceptual/Procedural Distinction. *DPhil Thesis* . Magdalen College, Oxford.
- Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases* . Amherst, Mass.: University of Massachusetts.
- Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.
- Holmberg, A. (2004). Null subjects and uninterpretable features: evidence from Finnish. *Durham Working Papers in Linguistics* 9 .
- Holmberg, A. (2005). Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* , 36 (4), 533-564.
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, Form and Use in Context: Linguistic Applications (GURT '84)* (pp. 389-418). Washington DC: Georgetown University Press.
- Horsey, R. (1998). Null Arguments in English Registers: a minimalist account. *BA Thesis* . Melbourne, Australia: La Trobe University.

- Huang, C.-T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15 (4), 531-574.
- Hulme, K. (1985 [1984]). *The Bone People*. London: Hodder & Stoughton.
- Hussein, M. (2008). The truth-conditional/non-truth-conditional and the conceptual/procedural distinctions revisited. *Newcastle Working Papers in Linguistics*, 14, 61-80.
- Hyams, N.M. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.
- Hyams, N. M. (1992). A reanalysis of null subjects in child language. In J. Weissenborn, H. Goodluck, & T. Roeper (Eds.), *Theoretical Issues in Language Acquisition: continuity and change in development* (pp. 249-267). Hillsdale NJ: Lawrence Erlbaum.
- Hyams, N. M. (1994). V2, Null Arguments and COMP Projections. In T. Hoekstra, & B. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (pp. 21-55). Amsterdam: Benjamins.
- Hyams, N. M. (1996). The underspecification of functional categories in early grammar. In H. Clahsen (Ed.), *Generative Perspectives on Language Acquisition* (pp. 91-128). Amsterdam: John Benjamins.
- Hyams, N., & Jaeggli, O. (1988). Morphological Uniformity and the Setting of the Null Subject Parameter. In J. Blevins, & J. Carter (Ed.), *NELS 18* (pp. 238-253). Amherst, Mass.: University of Massachusetts.
- Hyams, N., & Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguistic inquiry*, 24, 421-459.
- Ifantidou-Trouki, E. (1993). Sentential adverbs and relevance. *Lingua*, 90, 69-90.
- Ionin, T. (2006). This is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics*, 14 (2), 175-234.
- Iten, C. (2005). *Linguistic Meaning, Truth Conditions and Relevance: the case of Concessives* (Vol. 10). Basingstoke: Palgrave.
- Jaeggli, O., & Safir, K. (1989). *The Null Subject Parameter*. Dordrecht: Kluwer.
- Janssen, T. J. (1995). Deixis from a cognitive point of view. In E. Contini-Morava, & B. S. Goldberg (Eds.), *Meaning as Explanation: Advances in Linguistic Sign Theory* (pp. 245-270). Berlin and New York: Mouton de Gruyter.
- Jensen, B. (2003). Syntax and Semantics of Imperative Subjects. *Tromsø Working Papers on Language and Linguistics*, 31 (1), 150-164.
- Jespersen, O. (1940). *A Modern English Grammar Part V*. Copenhagen: Munksgaard.

- Joshi, A. K., & Weinstein, S. (1981). Control of inference: role of some aspects of discourse structure - centering. *Proceedings of 7th International Conference on Artificial Intelligence*, (pp. 385-387).
- Kaplan, D. (1989). *Themes from Kaplan*. Oxford: Oxford University Press.
- Kasher, A. (Ed.). (1998). *Pragmatics: Critical Concepts*. London: Routledge.
- Katz, J., & Postal, P. M. (1964). *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: M.I.T. Press.
- Kazman, R. (1988). Null arguments and the acquisition of Case and Infl. *13th Boston University Conference on Language Development*. Boston.
- Kemmerer, D. (1999). "Near" and "far" in language and perception. *Cognition*, 73 (1), 35-63.
- Lakoff, R. (1974). Remarks on This and That. *Papers from the Tenth Regional Meeting of the Chicago Linguistics Society*. Chicago: University of Chicago.
- Lasnik, H., & Stowell, T. (1991). Weakest cross-over. *Linguistic Inquiry*, 22, 687-720.
- Lee, M. S. (1993). The origin of the turtle body plan: bridging a famous morphological gap. *Science*, 261 (5129), 1716-1720.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47, 1940-1951.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (1998). Minimization and conversational inference. In A. Kasher (Ed.), *Pragmatics: Critical Concepts* (Vol. 4, pp. 545-612). London: Routledge.
- Levinson, S. C. (2000). *Presumptive Meanings: the Theory of Generalized Conversational Implicature*. Cambridge, Mass: MIT Press.
- Levinson, S. C. (2004). Deixis. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 97-121). Oxford: Blackwell.
- Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge: Cambridge University Press.
- Maclaren, R. (1980). On two asymmetrical uses of the demonstrative determiners in English. *Linguistics*, 18, 803-820.
- MacWhinney, B. (2000). *The CHILDES-project: tools for analysing talk*. (3rd Edition ed.). Hillsdale, NJ: Erlbaum.
- Maes, A. (1996). *Nominal Anaphors, Markedness, and the Coherence of Discourse*. Leuven: Peeters.
- Massam, D., & Roberge, Y. (1989). Recipe context null objects in English. *Linguistic Inquiry*, 20 (1), 134-139.
- Matsui, T. (2000). *Bridging and Relevance*. Amsterdam: John Benjamins.

- Matthews, A., & Chodorow, M. S. (1988). Pronoun resolution in two-clause sentences: effects of ambiguity, antecedent location and depth of embedding. *Journal of Memory and Language* , 27, 245-260.
- Mauck, S., Pak, M., Porter, P., & Zanuttini, R. (2005). Imperative subjects: A crosslinguistic perspective. *Georgetown University working papers in theoretical linguistics* , 135-152.
- Mazuka, R., Lust, B., Wakayama, T., & Snyder, W. (1986). Distinguishing effects of parameters in early syntax acquisition: a cross-linguistics study of Japanese and English. *Papers and Reports on Child Language Development* , 25, 73-82.
- McClelland, J. L. (1994). The organization of memory: a parallel distributed processing perspective. *Revue Neurologique (Paris)* , 150 (8-9), 570-579.
- McCloskey, J., & Hale, K. (1983). On the Syntax of Person-Number Inflection in Modern Irish. *Natural Language & Linguistic Theory* , 1 (4), 487-533.
- McKoon, G., Ratcliff, R., Ward, G., & Sproat, R. (1993). Syntactic prominence effects on discourse processes. *Journal of Memory and Language* , 32, 593-607.
- Meisel, J. (1987). Early differentiation of language in bilingual children. In K. Hyltenstam, & L. Obler (Eds.), *Bilingualism across the lifespan: in health and psychology* (pp. 13-40). Cambridge: Cambridge University Press.
- Napoli, D. J. (1982). Initial material deletion in English. *Glossa* , 16, 85-111.
- Neale, S. (2007). Heavy Hands, Magic, and Scene-Reading Traps. *European Journal of Analytic Philosophy* , 3 (2), 77-132.
- Neeleman, A., & Szendroi, K. (2007). Radical Pro Drop and the Morphology of Pronouns. *Linguistic Inquiry* , 38, 671-714.
- Neeleman, A., & Van de Koot, H. (2002). The configurational matrix. *Linguistic Inquiry* , 33 (4), 529-574.
- Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research* , 155-167.
- Noveck, I., & Sperber, D. (2007). The why and how of experimental pragmatics: the case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 184-212). Basingstoke: Palgrave MacMillan.
- Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy* , 3, 143-184.
- Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics* , 12, 109-132.
- Nunberg, G. (2004). The pragmatics of deferred interpretation. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 344-364). Oxford: Blackwell.

- Oh, S.-Y. (2006). English zero anaphora as an interactional resource II. *Discourse Studies* , 8, 817-846.
- Olson, G. (1973). Development changes in memory and acquisition of language. In T. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. London: Academic Press.
- Papafragou, A. (1996). On Generics. *UCL Working Papers in Linguistics* , 8, 165-198.
- Pepys, S. (1985). *The shorter Pepys*. (R. Latham, & W. Matthews, Eds.) London: Bell & Hyman Ltd.
- Pierce, A. (1992). Language acquisition and syntactic theory: a comparative analysis of French and English child grammars. Dordrecht: Kluwer.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin.
- Platzack, C., & Rosengren, I. (1998). On the subject of imperatives: A minimalist account of the imperative clause. *The Journal of Comparative Linguistics* , 1, 177-224.
- Powell, G. (1998). The deferred interpretation of indexicals and proper names. *UCL Working Papers in Linguistics* , 10, 1-32.
- Powell, G. (2001). Complex demonstratives. *UCL Working Papers in Linguistics* , 13, 44-73.
- Powell, G. (2002). Reference and Relevance. *PhD Thesis* . University College London.
- Predelli, S. (1998). I am not here now. *Analysis* , 58, 107-115.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-255). New York: Academic Press.
- Radford, A. (1996). Towards a structure-building model of acquisition. In H. Clahsen (Ed.), *Generative perspectives on language acquisition*. Amsterdam: John Benjamins.
- Reboul, A. (1997). What (if anything) is accessibility? A relevance-orientated criticism of Ariel's Accessibility Theory of referring expressions. In J. H. Connolly, R. M. Vismans, C. S. Butler, & R. A. Gatward (Eds.), *Discourse and Pragmatics in Functional Grammar* (pp. 91-108). Berlin/NY: Mouton de Gruyter.
- Reboul, A. (1998). A relevance theoretic approach to reference. *Acts of the Relevance Theory Workshop* (pp. 45-50). Luton: University of Luton.
- Reboul, A. (1999). Reference, agreement, evolving reference and the theory of mental representations. In M. Coene (Ed.), *Traiani Augusti vestigia pressa sequamur :studia I linguistica in honorem L. Tasmowki*, (pp. 601-616). Padova: Unipress.

- Reinhart, T. (1981). Two COMP positions. In A. Belletti, L. Brandi, & L. Rizzi (Eds.), *Theory of Markedness in Generative Grammar* (pp. 601-616). Pisa: Scuola Normale Superiore.
- Riemsdijk, H. v., & Williams, E. (1986). *Introduction to the Theory of Grammar*. Cambridge, Mass.: MIT Press.
- Rizzi, L. (1986). Null objects in Italian and the theory of pro. *Linguistic Inquiry*, 17, 501-558.
- Rizzi, L. (1994). Early null subjects and root null subjects. In T. Hoekstra, & B. Schwartz (Eds.), *Language Acquisition in Generative Grammar* (pp. 151-176). Amsterdam/Philadelphia: John Benjamins.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar: Handbook in Generative Syntax* (pp. 281-337). Dordrecht: Kluwer.
- Rizzi, L. (2005). On the Grammatical Basis of Language Development: A Case Study. In G. Cinque, & R. S. Kayne (Eds.), *Oxford Handbook of Comparative Syntax* (pp. 70-137). Oxford: Oxford University Press.
- Sauerland, U., & Elbourne, P. (2002). Total reconstruction, PF movement, and derivational order. *Linguistic Inquiry*, 33, 283-319.
- Schaeffer, J. (2005). Pragmatic and grammatical properties of subjects in children with specific language impairment. *UCLA Working Papers in Linguistics 13, Papers in Psycholinguistics*.
- Schiffer, S. R. (1972). *Meaning*. Oxford: Oxford University Press.
- Schmerling, S. (1972). Subjectless sentences and the notion of surface structure. In C. Corum, T. C. Smith, & A. Weiser (Ed.), *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society* (pp. 577-586). Chicago: University of Chicago Press.
- Schmerling, S. (1982). How imperatives are special, and how they aren't. In R. Schneider, K. Tuite, & R. Chametzky (Ed.), *Chicago Linguistic Society: parasession on nondeclaratives*, (pp. 202-218).
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272-281.
- Shipley, E. F., Smith, C. S., & Gleitman, L. R. (1969). A study in the acquisition of language: free responses to commands. *Language*, 45, 322-342.
- Sidelle, A. (1991). The answering machine paradox. *Canadian Journal of Philosophy*, 21, 525-539.
- Smart, E. (1991). *Necessary Secrets*. (A. Van Wart, Ed.) London: Paladin.

- Speas, M. (1994). Null arguments in a theory of economy of projection. (E. Benedicto, & J. Runner, Eds.) *UMass Occasional Papers in Linguistics* .
- Speas, M. (2006). Economy, agreement and the representation of null arguments. In P. Ackema, P. Brandt, M. Schoorlemmer, & F. Weerman (Eds.), *Argument and Agreement*. Oxford: Oxford University Press.
- Sperber, D. (1994). Understanding verbal understanding. In J. Khalifa (Ed.), *What is Intelligence?* (pp. 179-198). Cambridge: Cambridge University Press.
- Sperber, D., & Wilson, D. (1986/95). *Relevance: Communciation and Cognition* (Second edition (with postface) ed.). Oxford: Blackwell.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language* , 17 (1), 3-23.
- Sperber, D., & Wilson, D. (2008). A deflationary account of metaphor. In R. Gibbs (Ed.), *Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press.
- Strauss, S. (1993). Why this and that are not complete without it. *Proceedings of the 29th Regional Meeting of the Chicago Linguistic Society: i: The Main Session* (pp. 403-417). Chicago Linguistic Society.
- Strauss, S. (2002). This, That and It in Spoken American English: a demonstrative system of gradient focus. *Language Sciences* , 24 (2), 131-152.
- Stroumza, A. (1998). *Le Journal de Bridget Jones*. Paris: Albin Michek.
- Sweet, H. (1960). *New English Grammar*. London: Oxford University Press.
- Talmy, L. (1988). The relation of grammar to cognition. In B. Rudska-Ostyn (Ed.), *Topics in Cognitive Linguistics*. (pp. 165-206). Amsterdam: John Benjamins.
- Thorne, J. P. (1966). English Imperative Sentences. *Journal of Linguistics* , 2 (1), 69-78.
- Thrasher, R. (1977). One way to say more by saying less: a study of so-called subjectless sentences. *Kwansei Gakuin University Monograph Series 11* . Tokyo: Eihosha.
- Tomioka, S. (2003). The semantics of Japanese null pronouns and its cross-linguistic implications. In K. Schwabe, & S. Winkler (Eds.), *The Interfaces: Deriving and Interpreting Omitted Structures* (pp. 321-339). Amsterdam: John Benjamins.
- Valian, V. (1990). Null subjects: a problem for parameter-setting models of language acquisition. *Cognition* , 35 (2), 105-122.
- Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* , 40, 21-81.

- Valian, V., Aubry, S., & Hoeffner, J. (1996). Young children's imitation of sentence subjects: evidence of processing limitations. *Developmental Psychology*, 32 (1), 153-164.
- Walker, M. A., Joshi, A. K., & Prince, E. (Eds.). (1998). *Centering Theory in Discourse*. Oxford: Clarendon Press.
- Wexler, K. (1991). Optional infinitives, head movement and the economy of derivations in child grammar. *Occasional Paper 45, Center for Cognitive Science*.
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua*, 106 (1), 23-79.
- Wharton, T. (2003). Interjections, language, and the showing/saying continuum. *Pragmatics and Cognition*, 11 (1), 39-91.
- Wharton, T. (2009). *Pragmatics and Non-Verbal Communication*. Cambridge: Cambridge University Press.
- Wharton, T. (forthcoming). Beyond the words. *Gastronomica*.
- Wilson, D. (1992). Reference and relevance. *UCL Working Papers in Linguistics*, 4, 167-191.
- Wilson, D. (2000). Metarepresentation in Linguistic Communication. In D. Sperber (Ed.), *Meterepresentations* (pp. 411-448). Oxford: OUP.
- Wilson, D. (2009). The conceptual-procedural distinction: Past, present and future. Paper presented to the International Conference on Procedural Meaning, Madrid.
- Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, 90, 1-25.
- Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111 (443), 583-632.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.
- Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, 38 (10), 1559-1579.
- Wittek, A., & Tomasello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics*, 26 (4), 541-558.
- Woolf, V. (1985). *The diary of Virginia Woolf* (Vols. 5: 1936-1941). (A. McNeillie, & A. O. Bell, Eds.) Middlesex, England: Penguin.
- Zaki, M. (2009). Demonstratives and the Conceptual/Procedural Distinction. *Paper Presented at Procedural Meaning: problems and perspectives*. Madrid.
- Zwicky, A., & Pullum, G. (1983). Deleting named morphemes. *Lingua*, 59, 155-175.