Reinventing Pronoun Gender

VRIJE UNIVERSITEIT

Reinventing Pronoun Gender

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Letteren
op vrijdag 13 november 2009 om 10.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Jenny Audring

geboren te Berlijn, Duitsland

promotoren: prof.dr. G.E. Booij
                    prof.dr. G.G. Corbett

Acknowledgements

reinvent, *v.*

▶ *trans.* To adopt a new image or identity for (a person or thing). Usu. *refl.*: to adopt a new image or identity for oneself; to change one's behaviour in order to respond to a change in environment or react to opportunity.
(Oxford English Dictionary)

***

"l'étude du pronom est la partie la plus urgente des études grammaticales"
(Hjelmslev *Principes de grammaire générale*, 1928: 331)

Abbreviations and styles

Glossing follows the Leipzig Glossing Rules
(http://www.eva.mpg.de/lingua/resources/glossing-rules.php)

Other abbreviations and styles:

AFF	affirmative
ANS	Algemene Nederlandse Spraakkunst (Dutch grammar)
E-ANS	electronic edition of the Algemene Nederlandse Spraakkunst (Dutch grammar)
C	common gender
CGN	Corpus Gesproken Nederlands (Corpus of Spoken Dutch)
F	feminine gender
M	masculine gender
N	neuter gender
PRT	discourse particle
WALS	World Atlas of Language Structures (Haspelmath et al. 2008)

-	in examples: change of turn/speaker
'	marker for clitics (according to Dutch spelling conventions)
[M], [human]	in examples/in text: features (semantic or formal)
italics	in text: quoted words or word parts
bold italics	technical term on first use
boldface	highlighting (in text or examples)
'single quotes'	translation
SMALL CAPS	grammatical features

No responsibility can be assumed for the stability of internet links (URLs) given in the text.

Contents

Introduction

This book is an empirical investigation of pronominal gender agreement in modern spoken Dutch. Because of a mismatching inventory of noun and pronoun genders, pronominalization in Dutch is riddled with conflicts and variation. Identifying and interpreting the choice patterns is the first major task of this work.

With the help of corpus data, the study identifies a number of syntactic and semantic factors that determine pronoun preferences in Dutch. The semantic factors together form a typologically interesting pattern. Especially, the parameter [±countable] or *degree of individuation* that governs pronominalization for inanimate referents merits discussion. Evidence from other Germanic languages shows that this parameter is more widespread than may be expected.

Semantics-based pronoun choice is often in conflict with the still existing syntactic system of lexical gender. In the competition of the systems, patterns of variation can be shown that are of interest to typology.

If the development of a new semantic gender system is seen as a solution for the morphosyntactic mismatch problem, this coping strategy is interesting for wider typological issues, in particular for the question of what happens if gender is reduced to marking on pronouns. An explorative typology of pronominal gender languages supports the hypothesis that not every type of gender system can survive on such impoverished agreement.

The main approach is typological and empirical, and the book contains a multitude of examples. Especially the Dutch data is presented in great detail in order to illustrate the richness and systematicity of what many speakers consider a marginal, substandard kind of language use. Often seen as mistakes that signal the decline of the system, the new pronoun genders are argued to be an ingenious 'reinvention' that consolidates it.

Part I

Pronouns, Gender and Agreement

Chapter 1

Pronouns

Personal pronouns are a fascinating part of speech. Among the many questions one can ask about pronouns, two of the most popular are "When is a pronoun chosen instead of a noun?" and "How is the antecedent of a pronoun identified?" Especially the latter topic, often referred to as "pronoun resolution", figures largely in linguistic literature. This book is about a problem that occurs in between the first two fields of interest. Its central question is, informally put: "When a speaker has decided to use a pronoun, which pronoun does s/he choose and why?" This question is of course most interesting in situations when there is more than one candidate pronoun with different grammatical features to choose between. In particular, we will be looking at pronouns with a choice in gender. The patterns of variation provide interesting insights for our theories about gender and agreement.

1.1 Sorts

The primary object language of this study is modern spoken Dutch. The pronouns considered are clitics and free words that are used to introduce relative clauses (relative pronouns), indicate possession or attachment (possessive pronouns) or refer to a conceptual entity mentioned or otherwise entertained in the previous discourse (personal pronouns). Furthermore, demonstrative pronouns in anaphoric use are considered. An example for each sort of pronoun is given in (1).

(1) a) Relative pronoun
    Het   boek   **dat**   ik   lees
    DEF.N   book(N)   REL.N   I   read
    'The book that I'm reading'

   b) Possessive pronoun
    Het   kind   en   **zijn**   fiets
    DEF.N   child(N)   and   POSS.N   bike(C)
    'The child and its bike'

   c) Personal pronoun
    De   man   zag   dat   **hij**   alleen   was
    DEF.C   man(C)   saw   that   3.M   alone   was
    'The man saw that he was alone'

d) Demonstrative pronoun

Zij	zag	de	jongen	maar	**die**	keek	weg
3.F	saw	DEF.C	boy(C)	but	DEM.C	looked	away

'she saw the boy but he looked away'

The personal pronouns are the centre of attention, and the term *pronoun*, if used without specification, is meant to refer to this type. The central condition is that the pronouns in question agree in gender with their antecedent noun. Thus, amongst the personal pronouns, only the third person pronoun singular is of interest, as Dutch personal pronouns do not mark gender in the first and second person and the language has no gender distinctions at all in the plural.

Sometimes, determiners are referred to as pronouns in linguistic literature. Especially with regard to demonstratives, the distinction between anaphoric and attributive is not always made consistently. In the present study, the two domains need to be kept separate. The main motivation is that determiners and attributive demonstratives belong to a different agreement domain - viz. the noun phrase - than free pronouns and anaphoric demonstratives. More specifically, this study is interested in agreement **variation**, and Dutch, like most languages, does not show systematic variation in attributive agreement (i.e. within the NP). Thus, an anaphoric neuter pronoun such as *het* in (2b) is a relevant item for this study, while the homophonous *het* in (2a) is not. The same holds for the two demonstrative forms in (2c) and d). Only (2d) is relevant.

(2) a) **het** kind
DEF.N child(N)
'the child'

b) ik heb **het** gisteren gezien
I have 3.N yesterday seen
'I have seen it yesterday'

c) **die** man
DEM.C man(C)
'that man'

d) als je **die** ziet
if you DEM.C see
'if you see him'

Indefinite pronouns such as *iets* 'something', *iemand* 'somebody' or *elk* 'each' and question words such as *wie* 'who' or *wat* 'what' are not considered. Motivation for this choice is given in Chapter 4 on the methodology for the Dutch corpus study that constitutes the centrepiece of the present work.

1.2 Function: Anaphoric, exophoric and deictic

The pronouns discussed in this book are mainly personal pronouns functioning as anaphors. In the straightforward instances, an anaphoric pronoun has an antecedent, typically a noun phrase, with which it is coreferent. In less prototypical cases, the antecedent can be a clause, another pronoun or simply be absent from the utterance. Coreference is not necessarily watertight either. Often, antecedent and pronoun reflect two semantic construals that amount to two different entities in the world. We will meet relevant cases in Chapter 7.

Pronouns, however, can also be used deictically, as in (pointing) *He (there) is my new neighbour*. This usage is sometimes referred to as *exophoric* (e.g. Halliday and Hasan 1976, Diessel 1999), and some studies prefer the term *endophoric* for what is more commonly known as anaphoric reference (e.g. Cornish 1999). Deictic or exophoric pronouns are said to refer directly to entities in the world, without an antecedent in the linguistic context. Moreover, they often have the function of moving a new referent into the focus of attention (Bosch 1983: 56). By contrast, anaphoric elements can only be used when the intended referent is already salient in the discourse. Yet, in practice, the distinction is much less clear. This chapter is intended to review some of the difficulties.

In many languages, deictic and anaphoric pronouns are homophonous - or rather, the same pronouns can be used for both functions - and the different uses may appear in very similar linguistic environments. An example is the following hypothetical mini-dialogue discussing, for instance, a number of photographs on the table. (A and B are different speakers. Here, as in all examples, the relevant items are highlighted in boldface).

(3) Dutch
   A: Heb   je   **die**   al?   (pointing)
       Have you DEM.C   already

   B: Nee **die**   heb ik nog niet,
       No   DEM.C   have I   yet   not

       maar **die**   wel. (pointing)
       but   DEM.C   AFF

       'Have you got this one? - No, that one I haven't got, but this one I have.'

In the first utterance, *die* is used deictically: it has no antecedent and the (new) referent is singled out with the help of a pointing gesture. The same is true for the last *die* in the second utterance. The function of the middle *die*, though, is difficult to assess. Accompanied by a gesture, it could be deictic. Without gestural support, the pronoun qualifies as an anaphor, as it refers to an entity already in focus. Prosodic information such as stress or accent, which is often used as a diagnostic to

distinguish deictic from anaphoric pronouns (e.g. Bosch 1983: 59), is no help in this case. By virtue of the contrastive meaning, all three pronouns have the same stress value.

If the second pronoun is an anaphor, the question is which element is its antecedent. Clearly, the only candidate is the first, the deictic pronoun, and indeed, the two pronouns have the same feature values, singular number and common gender (they are indeed homophonous). However, the antecedent is not a nominal with independent lexical features of its own. It is a deictic pronoun whose features in turn must stem from some other source. In many languages, deictic pronouns match the grammatical features of the noun that would be used to refer to the entity pointed at (Tasmowski-De Ryck and Verluyten 1982, Tasmowski and Verluyten 1985). Thus, the German equivalent of the dialogue in (4) would contain neuter pronouns in agreement with the word *Foto* (N) 'photograph'.

(4)   German
        A: Hast   du   **das**   schon?   (pointing)
             Have   you   DEM.C   already

        B: Nein   **das**   habe   ich   noch   nicht
             No   DEM.C   have   I   yet   not

             aber   **das**   hier   schon   (pointing)
             but   DEM.C   here   AFF

             'Have you got this one? - No, that one I haven't got, but this one I have.'

Here, the deictic pronoun agrees with an implicit noun. Masculine or feminine pronouns would be ungrammatical. If this is the case, what prevents us from thinking that the subsequent anaphoric pronoun in the Dutch and the German example, rather than agreeing with the deictic pronoun, agrees with this implicit noun, too?

In fact, there are pronouns that, while clearly anaphoric, may fail to have an overt antecedent. Two cases are given in (5). (5a) is part of an actual utterance, found in the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN).[1]

(5)   a)   als   ze   **een** **bloemetje**   koopt   dan   zet   ze   **ze**
            if   3.F   a   flower.DIM(N)   buys   then   puts   3.F   them

---

[1] Examples from the Corpus Gesproken Nederlands (Corpus of Spoken Dutch) are marked as "CGN", followed by the number of the recording session. Recording sessions make up the internal structure of the corpus. Noting the session number should help to locate the examples in the corpus.

in	de	WC	want	dan	blijven	**ze**	langer	goed
in	DEF.C	wc(C)	because	then	stay	they	longer	good

'If she buys a flower then she puts them in the bathroom because then they will stay fresh longer.'
(CGN session 311)

b)
Ik	heb	**geen boeken**	behalve	**dit**
I	have	no book(N).PL	besides	3SG.N

'I've got no books besides this one.'

Here, the pronouns are anaphors, yet their antecedent is, strictly speaking, not overtly present. The nouns that are treated as antecedents do not have the same features as the pronoun - they bear a different number value - nor are they completely coreferent. In (5a), the noun refers to one flower but the pronoun to several, and in (5b), the noun indicates a number of unspecific books but the pronoun exactly and specifically one. Yet, if the pronouns are anaphoric and singular, and when the language is Dutch or, say, German, they need to agree in gender.

Another problematic sort of data are anaphoric pronouns that have no antecedent at all. Such pronoun usage is possible when linguistic context and/or human interaction narrow down the range of potential discourse referents to such an extent that only one is available for pronominal reference. (6) is a relevant example from the Corpus of Spoken Dutch. In (6), the verb *pinnen* 'to withdraw money from a cash machine' so strongly activates the concept of cash machines that this can be picked up by anaphoric pronouns.

(6)
weet	je	dat	je	tegenwoordig	ook	op	het
know	you	that	you	nowadays	also	on	def.N

station	heel	dicht	bij	ons	kunt	pinnen? -
station(N)	really	close	with	us	can	withdraw_money

ja	maar	**die**	zijn	nog	niet	in	gebruik.
yes	but	**they**	are	yet	not	in	use

'Did you know that you can now withdraw money at the station, really close to us? - Yes but they are not in use yet.'
(CGN session 446)

Here, the plural pronoun *die* unambiguously refers to cash machines, although these are not explicitly mentioned in the dialogue.

In another example, reference is established without any previous linguistic context. Two contributors to the corpus, aware of the recording device in their room, have the following interchange.

9

(7) A: doet **ie** 't wel goed?
         does 3.M 3.N PRT well
         'Is it working alright?'

    B: ja **hij** doet 't goed. is geen probleem.
         yes 3.M does it well is no problem(N)

         hij neemt gewoon op.
         3.M takes normally on

         'Yes it's working ok. No problem. It's recording normally.'
         (CGN session 392)

There is no overt noun phrase that could serve as the antecedent of the highlighted pronouns. Neither are any of the pronouns accented in a way that indicates the salience-shift typical for pronoun deixis (Bosch 1983: 59). Yet, speaker B reacts in a way that indicates he has understood A's question. This example is especially interesting as corpus participants are generally uncertain what to call the recorder - it is mostly referred to as *ding* [N] 'thing' - so it is unclear which noun could have triggered the masculine gender of the pronoun.

In both (6) and (7), we have a case of an antecedentless anaphor, which the standard theory does not account for.[2] The presence or absence of an overt antecedent with matching features is obviously not a failsafe diagnostic for distinguishing deictic and anaphoric pronoun use. This is important if one wishes to make a principled theoretical distinction between (deictic) reference and (anaphoric) agreement. The difference is neither straightforward nor unproblematic.

Worse still, there are cases where an element is simultaneously deictic and anaphoric. Consider the following English example from Huddleston and Pullum (2002: 1454).

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

Speaking from a place outside London, "there" can be deictic, but also anaphoric with "in London" as antecedent. As Lyons puts it: "[w]hether the pronoun is interpreted as having anaphoric or deictic reference (or both) would seem to depend primarily upon the context-of-utterance" (Lyons 1977: 661); the alternative "or both" clearly indicating that the answer may be inconclusive. The same view is expressed in Cornish (1999) who writes: "anaphora and deixis [...] are discourse

---

[2] An exception is Cornish (1986) and (1999), whose discourse-function theory of anaphoric reference allows for the non-linguistic context to move referents into the focus of attention, making them available for anaphoric reference despite the absence of an overt antecedent. Similar ideas are voiced under the name of Accessibility Theory (e.g. Ariel 2004).

functions which I do not regard as having an absolute 'either/or' status [...], it being perfectly possible for a given indexical expression to express both functions simultaneously" (Cornish 1999: 31-32).

Last but not least, there are practical limitations to assessing the status of pronouns. Non-speech information such as gestures must be available to the investigating linguist in order to identify and interpret deictic pronouns and to distinguish them from anaphoric homophones. Since the primary data of this study consists of transcribed audio-recordings, access to non-verbal information is precluded.

In order to avoid theoretical distinctions that lack solid evidence, the present study will treat all pronouns with overt antecedents as anaphors. Those pronouns that do not have an overt antecedent will be gathered under the term *exophoric*. This does not mean that they are deictic, only that their antecedent is implicit. In the empirical part of this study, exophoric pronouns will be mentioned when the linguistic context is sufficiently restricted as to allow the identification of the intended referent.

This choice is not intended to deny the existence or the theoretical importance of deictic pronouns. The deictic function constitutes a potential source for the semantic freedom of pronouns, a freedom that often makes them appear as more than mere function words. Yet, pronouns that have the 'wrong' features in syntax will never be dismissed as deictic and therefore non-agreeing. Such an explanation lacks principled ground unless deictic and anaphoric pronouns can be distinguished unambiguously, and unless we can be sure that deictic pronouns do not agree. As argued above, neither condition is met for the Dutch data.

Sometimes, pronouns are mentioned in connection with another type of deixis, known as "discourse deixis". Pronouns have a discourse deictic function when they refer to propositions rather than to entities (Himmelmann 1997, Diessel 1999). (9) is a constructed example.

(9) Hij   wilde   de   wekker   zetten maar is   **het** vergeten.
     3.M   wanted DEF.C alarm_clock(C) set   but   is   3.N   forgot
     'He had wanted to set the alarm clock, but forgot all about it'

Discourse deixis is only of marginal interest in this study as propositional antecedents are not nominal and therefore do not have gender.

1.3 Referentiality

The pronouns in this study are referential, i.e. they refer to a conceptual entity (a person, an animal, an object, a substance, an abstract notion, or a proposition). Non-referring pronouns, as in expletive constructions, are excluded (see section 5.3.1 for examples). Distinguishing referring from non-referring pronouns and establishing the exact referent is not always unproblematic. Chapter 4 discusses problems and methods.

1.4 Features

The main concern of this book is pronominal agreement. Thus, pronouns will be discussed as formally expressing morphosyntactic features. Pronouns normally agree in person, gender and/or number. In the present study, gender marking is of primary interest. The most important criterion is variance, i.e. some degree of freedom to choose between one form or another. This study is limited to third person pronouns, as gender marking is absent from first and second person pronouns in Dutch, the primary language considered.

1.5 Mismatch, variation, switch

Gender information can surface in various places in the sentence, and it is usually expected that all gender-marking elements that agree with the same noun will have the same gender value. This is a reflection of the current view on gender, which regards it as a system of agreement classes. In order for a noun to belong to a particular gender, it should consistently control a set of agreements on all targets. In some languages, however, the gender-indicating agreement targets differ among each other. These differences can lie in the range of genders their morphology enables them to mark, and in the gender values they come to bear, given a particular controller. This study is interested in languages whose pronouns have morphological forms for a larger range of gender values than is otherwise typical for that language. In other words, the gender paradigm of pronouns should be more extensive than that of determiners, adjectives and other agreement targets. Moreover, we are interested in cases where the pronoun bears a different gender than other targets agreeing with the same controller. Such cases are particularly rampant in Dutch.

Paradigmatically, Dutch has the morphological means to distinguish three or four genders on anaphoric pronouns (the number depends on the analysis), but only two genders on other agreement targets (determiners, adjectives, relative pronouns). Such a situation is referred to as **mismatch**. Saying that Dutch has a mismatching gender system is thus a statement about the paradigms, i.e. the form inventories of the relevant agreeing elements.

(10) Mismatch in the Dutch target genders

Agreement target	Determiners, adjectives, relative pronouns	Personal pronouns
Genders	*common* *neuter*	*masculine* *feminine* *neuter* *(common)*

Second, Dutch pronouns can take different genders for the same controller. That is to say, there is *variation* in pronoun usage. An example is (11), which can be continued as in (11a-c). All three variants are acceptable sentences in Dutch.[3]

(11) Ken je   zijn dochtertje?
      know you his daughter.DIM.N

    a) **Dat**   is al   zeven.
       DEM.N is already seven

    b) **Zij** is al   zeven.
       3.F is already seven

    c) **Die**   is al   zeven.
       DEM.C is already seven

    'Do you know his little daughter? She's seven already.'

Examples (11b) and c) show that such variation can lead to different gender values within an agreement chain, i.e. in a syntactic structure with several agreement targets for one particular controller. In (11b), the definite article reflects the neuter gender of the diminutive *dochtertje* 'little daughter', while the pronoun, which agrees with the same noun, is feminine. The change of gender value between the elements in a syntagma is referred to as a *switch*. Thus, (11b) contains a switch from neuter to feminine gender, and in (11c) the gender switches from neuter to common. The three domains of inconsistency in gender agreement are schematized in (12).

(12) Levels of inconsistency in agreement

Domain	Inconsistency
usage	variation
paradigm	mismatch
syntactic string	switch

These three phenomena, gender variation, gender mismatch, and gender switches, are the core issues of this book.

1.6 The structure of the book

The book is structured along the following lines. After a preparatory chapter about pronouns and agreement we will begin with a brief look at the history of the Dutch gender system to see how the mismatch situation arose. Since the loss of gender markers and the subsequent uncertainty about gender use has roused much

---

[3] While (11a) would be the preferred option in writing, (11b) and c) sound more natural in spoken language.

controversy and not a little emotion among Dutch linguists and the critical public, we will dwell a moment on the issues of awareness, attitude, avoidance and hypercorrection. The chapter will conclude with a positioning of the Dutch situation in a cross-linguistic context and with a note on analysis.

Chapter 4 elaborates on the methodology of the corpus study, from data source to coding. This is done in considerable detail in order to be transparent about the theoretical and practical choices made.

Chapter 5 centres on the question: what motivates a speaker of Dutch to choose a pronoun with the 'wrong' gender? The chapter shows how the individual pronouns - masculine, feminine, neuter and common - are distributed in spontaneous speech. All instances where pronominal agreement violates syntax are systematized according to semantic patterns. These patterns are combined in Chapter 6 to give a uniform account on the distribution of syntactically 'dis-agreeing' pronouns in spoken discourse.

Chapter 7 dwells on the issue of variation, giving examples for inter- as well as intra-speaker variation. It also provides a brief sketch of semantic agreement in written language, to complement the spoken corpus data. The chapter concludes with a note on referential ambiguity.

With a syntax-based and a semantics-based system of gender agreement existing side by side, the question arises how the two systems interact or interfere with each other. Can we predict under which conditions speakers are more likely to opt for semantic rather than syntactic agreement? Chapter 8 compares the frequency of each type of agreement with respect to a number of factors that may influence the choice.

Chapters 9 and 10 broaden the view to other languages. In Chapter 9, Germanic languages are discussed, while Chapter 10 lifts the issues to a more general level and considers data from pronominal gender languages anywhere in the world. The book concludes with some thoughts on the interaction of gender assignment and gender agreement.

1.7 A note on glossing

The book contains numerous examples. They often consist of entire utterances or stretches of dialogue, which makes full glossing unattractive for reasons of space and readability. Therefore, in the Dutch examples, only gender-bearing and gender-agreeing elements are glossed, as well as function words without an English equivalent. Glosses for number and case are only provided when helpful or necessary. The gender of each noun is treated and glossed as an inherent, lexical feature, in spite of the theoretical intricacies posed by the conflicting agreement evidence. The relevant problems are discussed in section 3.2. Examples are from Dutch unless otherwise indicated.

Chapter 2

Pronouns and Agreement

Much in this study rests on the assumption that pronouns are agreeing elements just as determiners or predicates are. This is not an uncontroversial stance. Many studies exclude pronoun-antecedent relations from agreement, or set them apart as a separate type. Doubts about the status of pronouns as agreement targets are strongest when the pronouns in question fail to match the feature values of their antecedents. However, the relevant typological literature shows that anaphoric pronouns share many properties with what is canonically assumed to be agreement, while many of their apparently special properties are also found in other agreeing elements. Thus, a boundary between anaphoric dependencies and (other) agreement relations lacks principled ground. After some clarification of terminology, we will review the debate briefly. The chapter concludes with an introduction to Corbett's *canonicity* approach, with the help of which the phenomena discussed in this book can be situated in the theoretical space that is called agreement.

2.1 Agreement: definition and terms[1]

Agreement is defined as "systematic covariance between a semantic or formal property of one element and a formal property of another" (Steele 1978: 610). The first element is referred to as the *agreement controller*, the second as the *agreement target*. The property in which the target covaries with the controller is called a *feature*. Typical features are person, number and - the star of the present study - gender. Features have certain *values* (say singular or plural, masculine or feminine) which are formally marked on the target (though not necessarily on the controller). Finally, agreement operates in a particular *domain*, for instance between subject and predicate.

Controller and target stand in an asymmetric relation to each other. This asymmetry has a formal and a semantic side. On the semantic side, the information in the agreement marking is relevant to the controller, not to the target. Thus, in (1) (example from Corbett 2006: 1), the singular marking on the verb pertains to the singularity of the agent, Mary, not of the event: she may be making pancakes more than once (in fact, this is exactly what the simple present suggests).

(1)   Mary   makes   pancakes

---

[1] The terminology is adopted from Corbett (2006). Since the focus of interest is on anaphoric pronouns, we will additionally use the term *antecedent* for the agreement-controlling noun.

According to Moravscik (1988), this makes agreement a case of "displaced information" (90): the number information belonging to the controller *Mary* is displaced onto the verb. As we will see later, pronominal agreement targets are special with regard to this point.

On the formal side, the target depends on the controller for its feature specification. This implies that changing the controller is expected to have repercussions for the target, but not vice versa. Examples (2) from Italian illustrate the situation. Changing the noun in a) and b) changes the agreements, but changing the agreeing elements in c) has no consequences for the noun.

(2) a) Italian
         un-a   altr-a   donna
         INDEF-SG.F   other-SG.F   woman.SG(F)
         'another woman'

     b) un   altr-o   uomo
         INDEF.SG.M   other-SG.M   man.SG(M)
         'another man'

     c) la   vecchi-a   donna
         DEF.SG.F   old-SG.F   woman.SG(F)
         'the old woman'

For gender, the value of the controller noun is normally lexically specified (although some nouns can have more than one gender). This does not hold for number; exceptions are singularia and pluralia tantum. Lexical features are characterized by the absence of choice. Thus, communicative intentions are normally unable to interfere with gender agreement. However, gender is a *direct feature* (Zwicky 1992) which is "associated directly with prototypical, or default, semantics" (as opposed to *indirect features* such as case, finiteness or declension which "are not so directly meaningful", Zwicky 1992: 378). This semanticity may cause conflicts and variation. Many languages allow for gender agreement values to deviate in favour of a semantically justified alternative. Such *semantic agreement* will figure largely in the present work and will be introduced next.

2.2 Semantic vs. syntactic agreement

Steele's definition of agreement encompasses two possibilities. An agreement target can express a formal property of its controller - such as its lexical gender - or a semantic property. The two options are often called agreement *ad formam* (also *formal, morphosyntactic, lexical* or *grammatical agreement*) and agreement *ad sensum* (also *notional, logical, pragmatic* or *referential agreement* or *synesis*). Most

common are the terms *semantic* and *syntactic* agreement, and they will be used throughout this book.[2]

The difference between semantic and syntactic agreement is visible in cases where the semantic and formal properties of the controller trigger different values on the agreeing target. A well-known example for such a situation is plural agreement with collective nouns, as is possible in many varieties of English (notably British English). An example is (3).

(3)   The committee were arriving

Here, the auxiliary expresses the semantic plurality rather than the formal singularity of the controller noun. Languages differ in the extent to which they allow semantic agreement, but the variation we find in the languages of the world is not unconstrained. The universal pattern that has been identified is captured in the *Agreement Hierarchy* (Corbett 1979, 1991, 2006). This hierarchy gives a ranking of four agreement targets that vary in syntactic distance to the controller.

(4)   Agreement Hierarchy (Corbett 1979, 1991, 2006)[3]

attributive > predicate > relative pronoun > personal pronoun

---

[2] Despite the choice for the more common terms, there are arguments in favour of the terms "lexical" vs. "referential" or "pragmatic" agreement (Dahl 2000). The use of a gender that diverges from the syntactic is often pragmatically or referentially motivated rather than based on the (lexical) semantics of the noun. This can be seen in instances where the actual agreement feature reflects a property of the referent that depends on the construal of the particular situation. Examples are nouns referring to animals, which are often ambiguous with regard to the sex-differentiability of the referent. Thus, *dog* denotes a canine of either sex (in which case English speakers use *it*) or an individual animal of male or female sex (in which case it can be referred to as *he* or *she*). In these cases, the semantics of the noun allows several construals, to which the agreement is sensitive. More examples can be found in agreement that is sensitive to countability, as will be discussed in detail below. Countability as a semantic property is construal-dependent: many nouns can take a count or a mass reading, depending on context and communicative intention. When agreement is count-mass sensitive, then the lexical semantics of the noun will not be sufficient to predict the agreements. Corbett (2006: 156 note 12) motivates his preference for the traditional terms with the argument that not only individual lexemes, but also constructional mismatches, such as conjoined noun phrases, should be captured in the account.
[3] Sometimes, e.g. in Corbett (1991=, the Agreement Hierarchy is given with the symbol "<". This assigns lowest ranking to the attributive and highest to the pronoun. Corbett (2006: 207, footnote 1) uses the alternative notation also adopted here in order to indicate decreasing canonicity from left to right.

There is robust cross-linguistic evidence that elements in the left part of the hierarchy are more likely to agree syntactically, i.e. with the formal properties of the controller, whereas targets to the right show an increasing likelihood of semantic agreement. The increase is monotonic, that is, without intervening decrease. Agreement constraints can be manifested in the grammar of a language as categorical rules or as mere preferences. We may find sharp cut-off points such as in English, where, for many speakers, semantic agreement is possible on all targets, but never for the attributive:

(5)　*These committee...

In other cases, there is a gradual scale of preference, such as documented for the Russian noun *vrač* 'doctor'. This noun can trigger masculine or feminine agreement on all four agreement targets, increasing in likelihood from the attributive to the personal pronoun along the hierarchy (see Corbett 2006: 210 f for an overview of the data).

Agreement preferences are visible both on sentence and on corpus level. In sentences with, say, a relative and a personal pronoun, semantic agreement will be more likely on the latter than on the former. On corpus level, there will be more personal pronouns than relative pronouns that agree semantically with their antecedent. The generalization is stronger on corpus level, as there may be individual sentences where a relative pronoun agrees semantically while other pronouns show syntactic agreement. In the sample of spoken Dutch used in the present study, the only example of this exceptional type is (6).

(6)　op　welk　　**huis**　　je　nou　had　gereageerd　ja. -
　　on　which.N　house(N)　you　now　have　reacted　　　yes

　　op　**die**…　die　straten　zeggen　me　niet　eens　iets.
　　on　DEM.C　the　streets　say　　me　not　even　something

　　die　　ene　　daar　bij　jou　in　de　　　buurt
　　DEM.C　one　　there　by　you　in　DEF. C　neighbourhood(C)

　　dat　　jullie　mam　niks　　vond
　　REL.N　your　mum　nothing　found

　　'Which house offer did you react to? - Well, on that one... the street names mean nothing to me. The one there in your neighbourhood that your mother didn't like.'
　　(CGN session 6798)

Furthermore, *distance effects* have been observed within types of agreement target. This again has sentence and corpus level application. When an utterance contains two personal pronouns, the one that is closer to the antecedent noun is more likely to agree syntactically, i.e. with the noun's formal or lexical properties. Similarly, the

overall distance between semantically agreeing pronouns and their antecedents will be greater than for syntactically agreeing pronouns. Again, individual sentences may show the reverse pattern. An example is (7).

(7) da   's   ook   al   een   oud   **programma**   **die**   is
     DEM.N is   also   already   a   old.N program(N)   DEM.N   is

     ook   later nooit geüpdatet. **'t**   werkt   ook   onder DOS
     also   later never updated   DEF.N   works   also under DOS

     'That's also an old program, it's never been updated later. It works under DOS, too.'

In Chapter 8, we will see that the cross-linguistic patterns are corroborated by the Dutch pronominalization data.

The fact that semantic agreement involves a controller and a target whose syntactic features do not match has prompted many linguists to set it apart from 'agreement proper'. Occasionally, the term "disagreement" is used. In formalist theory, different features are assumed for semantically and syntactically motivated agreement. Think, for example, of the three-way split between *concord*, *index agreement* and *pragmatic agreement* that is part of the framework of Head-Driven Phrase Structure Grammar (HPSG, for versions see Pollard and Sag 1994, Kathol 1999, Wechsler and Zlatić 1998). In particular, semantic agreement based on natural gender is often set aside as agreement in sex or indeed as direct reference to sex.

In the following, we will assume that both semantic and syntactic agreement constitute examples of gender agreement. The main reason for this choice is the belief that an element either is or is not an agreement target for a particular feature and a particular controller in a language. Thus, we say that definite articles in German agree with their noun in gender and number. This is a principal statement, which is valid even when, in a particular situation, the morphology fails to express the relation in the expected way. Else, whether or not an element is an agreement target becomes a case-by-case decision. Agreement relations are essentially grammatical and obligatory. Even when the speaker has a choice between alternative feature values, their distribution is not completely free, but subject to structural as well as semantic restrictions.

Whether or not a target and a controller are assumed to share an agreement relation is a choice informed by typological knowledge about the grammar of the language. If this knowledge gives reason to think that articles, adjectives or pronouns agree in gender, "[w]e need to see examples of mismatches against the background of all the cases where we do find matching" (Corbett 2006: 143). For Dutch and other Germanic languages, syntactic agreement is the norm for all of the three targets.

Separating agreement in gender from agreement in sex is rejected for other reasons. First, such a view disregards the fact that sex is one of the semantic cornerstones of

most gender systems (84 languages from the WALS sample (Haspelmath et al. 2008: 126 ff) have a sex-based gender system, against 28 gender languages that do not make this distinction). Moreover, 'agreement in sex' never seems to have its own distinct set of morphological markers (unless, of course, a language has a gender system purely based on natural gender, in which case the hypothesis cannot be tested). Rather, sex is expressed with the help of gender morphology. Only solid cross-linguistic evidence for sex agreement unrelated to gender would justify the analysis of sex as a separate morphosyntactic feature. Otherwise, 'agreement in sex' should be analyzed as semantic gender agreement.

The stronger version of the view, which says that pronouns refer to the sex of their referent directly rather than in their role as agreement targets, probably in much the same way that *man* and *woman* encode the distinction, needs some more discussion. It relates directly to the question if anaphoric pronouns are agreement targets at all. In the literature, pronouns are often given a special theoretical status, even when they do match their antecedent in gender. Yet again, the cross-linguistic evidence collected in connection with the Agreement Hierarchy shows that a strict division between anaphor-antecedent relations and other types of agreement is insufficiently motivated and typologically problematic.

2.3 Agreement vs. cross-reference

Antecedent-anaphor relations are frequently considered a separate phenomenon from phrase-internal or predicative agreement. They are often grouped under the term *cross-reference* or *anaphoric agreement* (Bresnan and Mchombo 1987). Corbett (2001 and 2006) gives a helpful overview of the different uses of the relevant terminology. The issue is treated in detail in Barlow (1991 and 1992: 134-152), Siewierska (2004: 221 ff), Corbett (1991, 2001 and 2006), and we shall briefly review the arguments here.

The choice to set anaphors apart seems to be motivated by at least three types of reason. The first reason is that of domain. Anaphoric pronouns are those elements that show the greatest syntactic freedom from their antecedent nouns. According to the nature of personal pronouns, which must be free in their syntactic domain, antecedent and anaphor are situated in different clauses.[4] By contrast, syntactic relations are often thought of as restricted to a local domain. In actual, especially spoken, discourse, pronouns can be widely separated from their antecedents, the distance stretching not only across utterances but even across turns in dialogue. Often enough, even intervening alternative antecedents do not break the chain. Apparently, pronouns are bound to their antecedents only by coreference, a semantic rather than a syntactic relation. These facts have prompted many linguists to set

---

[4] The syntactic freedom of pronouns is prominently expressed in Generative Grammar in Binding Principle B (Chomsky 1981) which says that pronouns must be free in their governing domain.

anaphoric pronouns apart from other agreement targets and discuss them under the heading of *cross-reference* (e.g. Bloomfield 1933).

What makes this split implausible is the fact that pronouns usually have the same features as other agreement targets and indeed the agreement-controlling nouns. On paradigm level, pronouns often show the same feature values as those of other agreeing elements. For example, in German, there are three different definite articles for masculine, feminine and neuter gender, as well as three different pronouns.

(8) Definite articles and pronouns in German

	Definite article	Personal pronoun
Masculine	*der*	*er*
Feminine	*die*	*sie*
Neuter	*das*	*es*

Similarly, in Russian, the past tense of the verb has markers for masculine, feminine and neuter, and there are three pronouns to match.

(9) Verbs and pronouns in Russian

	Verb (past tense) *prixodit'* 'to come'	Personal pronoun
Masculine	*prixodil*	*on*
Feminine	*prixodila*	*ona*
Neuter	*prixodilo*	*ono*

If cross-reference were a completely different phenomenon than agreement, it is not clear why pronouns should have the same gender values as other agreement targets.

On the syntagmatic side, a second reason for separating pronouns from other agreement targets is their liability to differ in feature values from their antecedent. This phenomenon has been discussed above under the name of *semantic agreement*. While pronouns may be particularly prone to this type of agreement, cross-linguistic research has shown that semantic agreement can be found on all other targets. The likelihood of semantic agreement increases gradually down the Agreement Hierarchy rather than sharply between other targets and the pronouns. The case of Russian *vrač* 'doctor' is particularly telling. For this noun, even attributive elements, the core domain of agreement for any theory, allow semantic agreement (Corbett 191: 184). At the other extreme, personal pronouns are able to agree syntactically. This fact is problematic for all theories that try to limit 'agreement proper' to noun-phrase- or clause-internal dependencies.

A third and related reason to set pronouns apart is that they may introduce new, i.e. non-redundant, information into the utterance. Thus, in the English sentence in (10), the pronoun reveals that the friend is female.

(10) If my **friend** goes to the cinema, I go with **her**.

The issue of redundancy is often invoked in descriptions of languages with participant marking on verbs which could be regarded as pronominal affixes or as verbal agreement markers (Bresnan and Mchombo 1987, Baker 1996, Evans 1999, Corbett 2003b, Mithun 2003). Yet, if it is promoted to a diagnostic for distinguishing agreement from other phenomena, many otherwise straightforward cases become problematic. A prime example is verbal person, number and/or gender agreement in pro-drop languages. In the Spanish example in (11), the verb marks first person plural, no matter if the agreement-controlling pronoun is present or not. To say that the verb agrees when the pronoun is present but does not agree when the pronoun is dropped, introduces a very artificial distinction between two nearly identical constructions.

(11) Spanish
    (Nosotros) pone-**mos** la mesa
    2PL   set-1PL.PRS DEF.SG.F table.SG(F)
    'We set the table'

The same holds for cases where the controller is present but lacks the necessary feature specification. In Russian verbs in the past tense, there is an agreement marker for gender, which has no counterpart in first and second person pronouns (compare (12a) and b)).

(12) Russian
    a) On   stoja-**l**   u okn-a
           3SG.M stand-PST.SG.M at window-N.GEN
           'He stood at the window'

    b) Ja   stoja-**la**   u okn-a
           1SG stand-PST.SG.F at window-N.GEN
           'I stood at the window'

In (12b), the verb technically speaking introduces new information, yet we would like to maintain that it agrees with the subject, just as in (12a). If, however, a theory wishes to place a condition of redundancy on agreement, it should be applied to all agreement targets, rather than just to the pronouns.

The position taken in this book is to treat all anaphoric pronouns as agreeing elements. In line with the research on semantic agreement, we will talk of agreement even if the controller and the pronominal target show mismatches in their feature values. Such cases will figure prominently in this book, which is, after all, a study on mismatches and variation.

There is ample evidence that Dutch pronouns qualify as agreement targets in much the same way as the pronouns of Russian or German. The best proof here are pairs of synonyms that differ in gender. When a noun is replaced by a synonym, the

referent stays the same, but we see the gender-related covariance in the morphology of the agreeing elements. The pronouns behave exactly as other agreement targets (compare the change of the definite article and of the pronoun).[5]

(13) a) **Dit   fototoestel** is niet van mij, **het** is van mijn broer.
         DEM.N camera(N)   is not of   me 3.N is of   my   brother

    b) **Deze   camera**   is niet van mij, **die** is van mijn broer.
         DEM.C camera(C) is not of   me 3.C is van my   brother

       'This camera is not mine, it's my brother's'

Additional motivation can be found in contrasting anaphoric pronouns with full noun phrase anaphors such as (14).

(14) **John** is late again. That **idiot** is never on time.

In Dutch, full NP anaphors can occur in gender combinations that would not be possible with pronominal anaphors. Examples are the antecedent-anaphor pairs in (15). The common gender noun *camera* 'camera' can take a neuter gender NP anaphor, but never a neuter pronoun. Conversely, the neuter noun *bier* 'beer' readily combines with a common gender anaphor such as *troep* 'stuff', but not with a common gender pronoun.

(15) a) **De   camera**   is nieuw, maar **dat   ding**   werkt niet
         DEF.C camera(C) is new   but   DEM.N thing(N) works no
         'The camera is new, but the thing doesn't work'

    b) **De   camera**   is nieuw, maar ***het** werkt niet
         DEF.C camera(C) is new   but   3.N   works no
         'The camera is new, but it doesn't work'

    c) **Bier**   is niet duur,   maar ik lust **die   troep**   niet
         beer(N) is not expensive but   I like DEM.C stuff(C) not
         'Beer is not expensive, but I don't like the stuff'

    d) **Bier**   is niet duur,   maar ik lust ***die** niet
         beer(N) is not expensive but   I like 3.C   not
         'Beer is not expensive, but I don't like it.'

---

[5] The pronoun in (13b) is a demonstrative, which has a common and a neuter form and as such has the morphology to agree syntactically with a common gender noun. In these examples, we neglect other pronoun forms that are possible in these utterances. They are the concern of later chapters.

Generally speaking, an antecedent of gender X does not prefer an NP anaphor with the same gender above one with a different gender. This means there is no systematic covariation in gender between an NP anaphor and its antecedent. This fact leads us to assume that, unlike pronominal anaphors, NP anaphors do not agree.

However, despite the general choice for including pronouns among other agreement targets, there are reasons to set pronouns apart. One of them is relevant for our purposes and deserves a moment's attention. As mentioned earlier, agreement normally involves the displacement of information (Moravcsik 1988: 90), such that "one word can carry the grammatical meaning relevant to another" (Corbett 2006: 1). Pronouns, however, are coreferent with their antecedent and as such share its values both semantically and morphosyntactically. Thus, they differ from other agreeing elements in that the displaced information they mark is also directly relevant to them. In (16a), the singular marking on the verb does not imply that the event occurs only once. In (16b), however, the singular marking on the pronoun implies that the referent of that pronoun is a singular entity. Moreover, the feminine gender of *she* reflects not only the gender of the referent of the antecedent, but also that of the referent of the pronoun, in both cases Mary.

(16)   a)   **Mary makes** pancakes
       b)   **Mary** makes pancakes when **she's** in a good mood.

This is because pronouns are nominal elements, coreferent and agreeing with other nominal elements. As such, their gender and number markers can be seen as the overt expression of their own referent's features rather than being merely displaced information belonging to another element. Speakers may then perceive the gender marking on a pronoun as conflicting with the semantics of the pronominal referent. This can be witnessed in German, where a few loanwords denoting female persons have masculine gender. An example is the Anglicism *Vamp* (M) 'vamp, seductive woman' (listed as masculine in Görlach 2001, although some speakers regard it as a neuter noun). A speaker of German may not mind that *Vamp* takes a masculine article, but a masculine pronoun is out of the question. Obviously, the feature masculine is felt to be in conflict with the semantics of the referent. This is only the case in pronominalization. The same holds for *Babysitter*, *Teenager* and *Cheerleader*, which take masculine articles and adjectives, but never masculine pronouns when the referent is female. That the issue here is one of semantic conflict can be seen from a comparison with neuter nouns referring to persons such as *Opfer* (N) 'victim'. Here, neuter anaphoric pronouns are acceptable (even if often not preferred). Apparently, the perceived conflict is less strong for neuter nouns: the neuter does not carry strong associations with natural gender.

However, the degree of semantic fit of the genders differs between languages and so does the propensity of the agreement targets to take semantic or syntactic agreement. We will thus abide by the conclusion that pronouns agree in much the same way as other agreement targets do.

2.4 The canonicity approach to agreement

Agreement is a diverse phenomenon, and theories differ in their treatment of its varied forms of appearance. A useful framework that avoids terminological confusion is Corbett's *canonicity* approach (Corbett 2003a, 2005, 2006, 2007). Here the different parameters involved in agreement are multiplied out to give a conceptual space in which an ideal point, the most canonical situation, is defined. All observed agreement phenomena can be arranged according to their divergence from this point. This approach offers a clear measure of differences without introducing artificial separations that may be difficult to motivate or to defend in a cross-linguistic context.

2.4.1 Canonical gender agreement

Examples of the canonical situation in gender agreement are the Spanish phrases in (17).

(17) Spanish
      a)   un-a   muchach-a   pequeñ-a
           INDEF-SG.F   girl-SG.F   little-SG.F
           'a little girl'

      b)   un-as   muchach-as pequeñ-as
           INDEF-PL.F   girl-PL.F   little-PL.F
           'a few little girls'

These examples are canonical in the following ways (based on Corbett 2006: 9).

The **controller**, the noun, is present, it marks gender (and number) overtly, and it is consistent in the agreement it takes. That is, *muchacha* only triggers feminine agreement on any target.[6]

The **targets** have bound expression of agreement (inflectional suffixes). They mark agreement obligatorily and productively. They double the marking on the noun.[7]

---

[6] The list of canonical properties with regard to the controller is completed by the point "part of speech is not relevant", which means that "given a domain, for instance, subject-predicate agreement, in the canonical case we do not need further information on the part of speech of the controller" (Corbett 2006: 12), e.g. if the subject is a noun or a pronoun. The same condition can be made for the agreement target. For both controller and target, this condition is vacuously fulfilled in our example (the examples illustrate particular controllers and targets).
[7] This is to distinguish it from elements that only mark agreement when the controller is absent, examples in Corbett (2006: 17 f).

The marking is regular, i.e. not suppletive, and alliterative in form. There is only a single controller for the targets. [8]

With regard to **domains**, the examples are canonical in that the agreement is asymmetric. This notion has been mentioned above: the gender of the article and the adjective depends on that of the noun. The agreement takes place in the local domain of the noun phrase. This domain is one of several where Spanish marks gender agreement.

The agreement **features** are number and gender, gender being the more canonical one by virtue of its being a lexical feature. The feature values of noun, article and adjective match, and there is no choice of values: for each of the two controllers, the targets can only have those particular gender values.

Finally, there are no **conditions** on the agreement relation. A very common condition on gender agreement is number: German gender agreement only works with singular controllers. However, the Spanish examples show gender agreement in the singular and the plural.

2.4.2 Non-canonical gender agreement

Pronominal gender agreement in spoken Dutch diverges in many respects from the canonical situation.

First of all, Dutch nouns do not mark gender overtly. While there are many derivational suffixes that correlate with a particular gender, they cannot be regarded as gender markers. Such a scenario is disallowed by our conceptualization of gender as a system of agreement classes and by the principle of Lexical Integrity (e.g. DiSciullo and Williams 1987) which does not permit agreement to be directly sensitive to a particular piece of bound morphology.

As a matter of difficulty, Dutch nouns do not only lack overt gender marking, but also consistent agreement behaviour: many nouns trigger different agreement values on different targets. An example is (18).

(18) mijn **broertje**   **die**   was   toen was **'ie**   nog   klein
      my   brother-DIM(N) DEM. C   was   then was 3SG.M   still small

    'my little brother he was then... was he still small'
    (CGN session 684)

The neuter gender noun *broertje* 'little brother' here triggers first common gender and then masculine gender agreement on the personal pronouns that are coreferent

---

[8] This condition caters for a few rare cases of "trigger-happy" agreement (Comrie 2003) where the target can have a choice between two potential controllers.

with it. (If there had been a definite article or an adjective, it would have showed neuter gender.) Such syntagmatic inconsistency has been introduced as *switch* in Chapter 1.

This leads us directly to the target properties. Dutch personal pronouns are non-canonical agreement targets in that they mark gender suppletively. More in line with the canonical situation, gender marking is productive and obligatory, and there is only one potential controller for each pronoun.

As regards domain, the pronouns stand in an asymmetric relation to their nouns: they depend in gender on the antecedent. This is the canonical situation. Also, they are one of several gender-agreeing elements, so the domain is one of multiple domains. Yet, Dutch has two genders that are only marked pronominally: the masculine and the feminine. Both genders have no other exponents: they do not surface on other agreement targets. These two genders represent a non-canonical situation because their markers are not members of a set, but are the sole means of expression for this gender. Moreover, antecedent-anaphor relations represent the least canonical domain. Pronouns stand outside the clause that contains their antecedent, and often outside the sentence or the turn as well. In spoken discourse, the distance between controller and target can be astonishingly large: the greatest distance observed in the present study is 328 words between the two (CGN session 628, the antecedent is *vriendje* 'boyfriend').

As regards features, Dutch pronouns agree in number and gender. Gender again is the more canonical because it is a lexical feature. However, one of the most apparent diversions from canonicity is that in many Dutch utterances the feature values of noun and pronoun do not match. Rather, many pronouns offer a choice between two or more gender values.

Finally, Dutch gender agreement only works for the singular. Thus, there is a number condition on the gender agreement in this language.

2.4.3 Paradigmatic mismatches

In addition to the parameters listed so far, the Dutch data is non-canonical in a paradigmatic dimension: it shows a mismatch in value sets between different targets. While the articles, adjectives, demonstratives and relative pronouns only offer a choice between two gender values, the personal pronouns possess the morphological means to distinguish three different genders. The situation can be schematized as in (19).

(19) Paradigmatic mismatch between targets

Target category:	Articles	Pronouns
Gender values:	*common*	*feminine*
		masculine
	neuter	*neuter*

This scenario is particularly problematic if we assume that the articles reflect the gender of the noun. Then the mismatch is not only among targets, but between controller and target.

Such mismatches are found in other languages of the world. They can be of three types. In the first type, the controller has more values than the target. This is the case in Hebrew. According to Corbett (2000: 95 ff and 2006: 145), some Hebrew nouns distinguish three numbers - singular, dual and plural - while the verbs that agree with them only mark singular or plural. The solution to such a mismatch is easy: for the dual nouns, plural verb agreement is used. Thus, two controller numbers map on one target number. An example is (20) (from Corbett 2006: 145).

(20)   Hebrew
    ha-yom-**ayim**   'avr-**u**   maher
    DEF-day-DU   pass.PST-3PL   quickly
    'the two days passed quickly'

The reverse type of mismatch, a target with more distinctions than its controller, is more problematic. After all, a controller feature cannot easily be mapped on two different target features. A relevant case is Inari Sami, where the verb has a dual marker not found on the nouns (examples from Corbett 2006: 146).

(21)   Inari Sami
    Almai   **kuálášt**   onne.
    man.SG.NOM.N   fish.3SG   today
    'The man is fishing today.'

    Alma-**h**   kuá'lášt-**ava**   onne.
    man-PL.NOM   fish-3DU   today
    'Two men are fishing today.'

    Alma-h   kuá'lášt-**eh**   onne.
    man-PL.NOM   fish-3PL   today
    'The men are fishing today.'

For dual referents, the noun bears plural marking. Thus, the mismatch is resolved with recourse to semantics, which is straightforward in number.

When the target with a higher number of values is a pronoun, the controller-target mismatch goes in both directions. As a target, the pronoun has more feature values than its controller, the noun. However, it can itself serve as agreement controller, for example to a following relative pronoun, which in turn may have the lower number of values. Dutch is a case in point (constructed example (22)).

(22) Haar **broer** ging niet mee. **Hij**, **die** zo gezeurd had...
 her brother(C) went not with 3.M who.C so nagged had
 'her brother wasn't coming along. He, who'd been nagging like that...'

Here the masculine gender pronoun agrees with a common gender noun. The pronoun is then followed by a relative pronoun that only has the choices common and neuter. It agrees with the noun and bears common gender marking.

The Dutch mismatch stands out for another reason. While both in Hebrew and in Inari Sami the mismatch and the related mapping problem can be solved by means of semantics, because the feature in question is the highly semantic feature number, such a strategy is less readily available for the gender mismatch in Dutch. In the example (22) above, the pronoun refers to a male person, so its masculine gender comes as no surprise. Natural gender belongs to the semantic core of many gender systems in the languages of the world, and it seems natural for speakers to resort to this property when they have to choose a pronoun. However, when the referent is inanimate, it is much less clear which pronoun should go with a common gender noun. After all, the gender system of Dutch is not primarily organized according to semantic principles and there are but a few obvious gender rules applying to inanimate nouns. Chapter 6 of the present work is devoted to the question how, in view of the low semanticity of Dutch gender, speakers choose a pronoun from the several options that the syntax, or rather the mismatched paradigms, leaves open.

2.5 Conclusion

In the first part of this book, the theoretical field was outlined in which the book is set. After an introduction to anaphors and deixis, we discussed the issue how anaphoric pronouns might be treated in a theory of morphosyntax. The position was defended that pronouns are agreement targets much like adjectives or predicates. This stance is taken even in cases where the features of pronouns and their antecedents fail to match. Thus, agreement variation is discussed in the same terms as the syntactic congruence normally expected.

The gender system of Modern Dutch was described as a non-canonical system, in which mismatches and variation abound. This calls for a closer look at the data from this interesting language.

Part II

The Dutch Pronominal Problem

Chapter 3

The Dutch Situation

Dutch has one of the most curious gender systems in Germanic. Due to the mismatch and variation described in the previous chapters, the basic questions "How many genders does the language have?" and "To which gender does noun X belong?" require different answers for articles, relative pronouns and personal pronouns. This situation is a consequence of historical changes in the paradigms of attributive modifiers and relative pronouns, which occurred around the Middle Dutch period. This chapter gives a brief outline of the developments and their consequences for the Dutch gender system. It concludes with a situationing of the Dutch problem in the cross-linguistic context and with an outlook at the analytical difficulties.

3.1 From the past to the present: The history of Dutch gender

3.1.1 Normativity and language planning

The gender mismatch in Dutch is a consequence of a historical development known as *deflection* which affects the agreement morphology of the language. Until the Middle Dutch period (12th to 15th century), Dutch had the traditional three-gender system that is common for Indo-European languages. Before around 1500, the difference between masculine, feminine and neuter gender was visible on adjectives and determiners, and - via declension classes - even on the nouns themselves. By progressive erosion of the markers and increasing syncretism, masculine and feminine became formally indistinct, giving us the contemporary pattern of common and neuter gender. However, the time scale and order of the changes are hard to retrace because of the heavy normativity that clouds most linguistic writings on this issue through the centuries.

The history of the Dutch gender system is a history of language pruning and planning. The early accounts, such as the *Twe-spraack van de Nederduitsche letterkunst* (1584, henceforth *Twe-spraack*), were firmly rooted in the Latin grammar tradition, which dictated six cases and three genders.[1] More interesting is

---

[1] Even in later centuries, it was not uncommon for a grammar to identify an ablative in Dutch. The encyclopedic work of Ten Kate (1723) goes as far as mentioning four variants: ablativus commerativus (*Ten zynen huize* 'to his house'), instrumentalis (*Dit wierd volbragt door/met hem* 'this was accomplished by him'), narrativus (*Men spreekt van hem* 'they speak of him') and discretivus (*Uit het huis* 'from/out of the house') (Ten Kate, Noordegraaf en Van der Wal 2001 [1723]: I 325).

the description of the Dutch vernacular in the Latin school grammar *Exercitium Puerorum*, which notes as early as 1485 the syncretism of nominative and accusative (Van der Wal 1988: 246) Generally, however, grammars up to the 19th century were largely prescriptivistic and aimed at enrichment and improvement of the native language on the basis of the classical languages Latin and Greek. In particular, efforts targeted the inflectional morphology, especially case and gender, equating rich inflections with a high state of development. Improvement efforts involved the promotion of individual dialectal variants to standard forms, the reconstruction of earlier forms, as well as the invention of new distinctions.[2] Inspiration was taken from the writings of highly valued authors from the past (see section 3.1.3) and the more conservative Flemish and Brabantian dialects in the south (Geerts 1966, but Van der Sijs 2004: 442). For the modern linguist, this means that neither the historical grammars nor the preserved literary texts can be trusted to provide realistic evidence about the colloquial language at a particular point in time.[3]

Moreover, what is now the Dutch language area used to be a conglomerate of quite heterogeneous spoken varieties whose documentation was not taken into consideration by the research agenda until recently. In the last decades, a few unbiasedly descriptive studies have appeared, e.g. by Van Leuvensteijn (1986, 1992, 1997, Van Leuvensteijn and Dekker 1990), comparing 16th century diaries from Gouda in Holland and Brugge in West Flanders, as well as Hogenhout-Mulder and Van Reenen (1988), a corpus study of 14th century Gronings (the dialect of Groningen, now the northernmost province of the Netherlands). Generally, however, there is no comprehensive account of the diachronic facts. We can only speculate up to what period in time the three-gender system was alive in the various regional and dialectal varieties or in the developing standard language, what form its morphological exponents took and by which route it was replaced by the two-gender system of today. Yet, a short sketch should be attempted. The following section will give a rough account of what Middle Dutch gender morphology looked like, focusing on those properties of the paradigms that were to pave the way from a three- to a two-gender system.

---

[2] Some of these inventions survive to this day. One of the most prominent - debated among the educated public - is the distinction between *hun* (3rd person plural dative) and *hen* (3rd person plural accusative). This artificial functional split of two dialectal variants dates back to the grammar of Christiaen van Heule (1625) and it is still propagated by style manuals, although only a minority of writers manages to adhere to the rules consistently (Van der Sijs 2004: 478 ff, E-ANS § 5·2·5·2·3).

[3] Revealing hints can be found in statements such as the following from the grammar of Van Heule (1625) which gives a list of some 1500 nouns and their genders, but concedes that "Dit onderscheid der geslachten en behouft in den rijm altijt niet nagevolgt te worden, want om die oorzaeke zouden de Rijmers al te nouw gebonden zijn" ('this difference of the genders does not always have to be followed in rhyming, because this would constrain the rhymers too strictly', 1625: 16). Kollewijn rightly interprets this statement as evidence for the artificiality of the gender distinctions expected in the written language (1916: 49).

3.1.2 Middle Dutch gender morphology

In Middle Dutch, noun phrases show differences between three genders on the determiner, the adjective and on the noun itself. Table (1), from Mooijaart and Van der Wal (2008: 40), gives the paradigms for the typical Middle Dutch noun phrase (abstracting from regional and dialectal differences). The nouns are *gast* [M] 'guest', *mensche* [M] 'man, human', *daet* [F] 'deed', *siele* [F] 'soul', *hof* [N] 'garden, farm' and *herte* [N] 'heart'. The examples were chosen in order to represent weak and strong declension (distinguished by the presence respectively absence of noun-final schwa in the nominative).

(1) Middle Dutch NP declension, singular

Case	Gender	Definite article 'the'	Adjective 'good'	Noun (strong declension)	Noun (weak declension)
Nominative	M	*die*	*goede*	*gast*	*mensche*
	F	*die*	*goede*	*daet*	*siele*
	N	*dat*	*goede*	*hof*	*herte*
Genitive	M	*dies/des*	*goets/goeden*	*gast(e)s*	*menschen*
	F	*dier/der*	*goeder*	*daet/dade*	*siele(n)*
	N	*dies/des*	*goets/goeden*	*hoves*	*herten*
Dative	M	*dien/den*	*goeden*	*gaste*	*mensche*
	F	*dier/der*	*goeder*	*daden*	*siele(n)*
	N	*dien/den*	*goeden*	*hove*	*herte*
Accusative	M	*dien/den*	*goeden*	*gast*	*mensche*
	F	*die*	*goede*	*dade*	*siele*
	N	*dat*	*goede*	*hof*	*herte*

It is obvious that late medieval Dutch already failed to provide distinct forms for many cells in the paradigm. For example, the definite article is the same for masculine and feminine in the nominative, and the adjectives have only four different forms across the twelve paradigm cells. Moreover, most of the case and gender markers on the noun, the determiner and the adjective contained or consisted entirely of /n/ or /ə/, two sounds that were very vulnerable for apocope. This means that the Middle Dutch paradigms, already riddled with syncretism, were reduced further by phonological processes stripping many elements of their case and gender markers.

Where the word-final /n/ did occur, its distribution has been associated with different functions. It could be a marker for accusative singular, as in (2a) vs. b).

(2)   a)   die   goed-e   knecht
           DEF.NOM good-NOM servant.NOM(M)

      b)   die-n   goed-en   knecht
           DEF.ACC good-ACC servantACC(M)

For the authors of the *Twe-spraack* (1584), the presence or absence of the *-n*-suffix was entirely a matter of case, not of gender: "de *-n*-vorm is voor hem in verband met het genus funktioneel indifferent" ('for him, the *-n*-form is functionally indifferent in relation to gender', Geerts 1966: 61, author's translation). In fact, the author of the *Twe-spraak* does not distinguish between masculine and feminine articles at all, and in his own writing he uses both *de* and *den* for masculine and feminine nouns, although *den* occurs by preference for masculines (Dibbets 1995: 58).

In other cases the *-n*-suffix was interpreted (or indeed installed) as a masculine form, as in the opposition between *d(i)e-n man* 'the-NOM man' vs. *d(i)e vrouw* 'the-NOM woman'. The linguist Van Hoogstraten, one of the most influential individuals in the history of Dutch gender, acknowledges this option, although he rejects it as confusing:

> eenigen willen [...] dat *den* ook somtyts zou kunnen staen voor den *noemer*, of eersten naemval, als *den Haeg, den oorlog, den oever, den aers*. Maer liever dan zulk eene wyde deure open te zetten voor de verwarringe, myde ik dezen naemval

> ('some people want *den* to sometimes stand for the nominative, or first case, as in 'the Hague', 'the war', 'the shore', 'the bottom'. But rather than open the doors to confusion, I avoid this case' (Preface to *Aenmerkingen over de geslachten der zelfstandige naemwoorden* (1700), annotated edition by De Bonth and Dibbets 1995: 8; author's translation)

He and other grammarians preferred *d(i)en* as a unique form for the accusative, though only for the masculine. Thus, the suffix *-n* became specialized for case **and** gender. Unfortunately, in varieties where the *-n* existed, it was often phonologically conditioned and only occurred before vowels or certain consonants (van Gestel et al. 1992: 67, Van der Sijs 2004: 437, cf. examples of Van Hoogstraten in the above quote). As such, it was an unreliable indicator for any grammatical feature, be it case or gender.

The same difficulties arise with stem-final /ə/ on nominals. In the Middle Ages, the presence or absence of schwa stood for what was referred to as the weak and the strong declension: weak declension nouns end in schwa, strong declension nouns in a consonant (remember Table (1) above). Declension class only vaguely correlated with gender. Weak declension nouns could be masculine (*cnape* 'boy'), feminine (*tonghe* 'tongue') or neuter (*bedde* 'bed', examples from Van der Sijs 2004: 428). As the pronunciation of word-final schwa became less common, the majority of

nouns still having it happened to be feminines. Consequently, it was reinterpreted as a marker of feminine gender. The association was strong enough to trigger changes in individual words: some other feminine nouns took on /ə/ (e.g. *dade* 'deed' and *helfte* 'half', originally *daet* and *helft*), while some masculines and neuters ending in /ə/ moved to feminine gender (e.g. *mane* 'moon', previously masculine, and *kinne* 'chin', previously neuter). Further changes weakened this correlation, too. Word-final /ə/ (respectively <e>, since the relevant literature is concerned with written language) also served as the dative case marker for masculine and neuter nouns of the strong declension. The grammarian Leupenius (Caron 1958: 32 (modern edition), Van der Sijs 2004: 429) extended this function to nouns of any gender, thus reanalyzing the element as a case rather than a gender marker. Yet, already in Late Middle Dutch, the dative case markers on masculine and neuter nouns were often dropped (Van der Wal, personal communication).

These and similar facts about varying distribution and phonological reduction of markers show how syncretism progressively removed the formal distinctions between the cases as well as between masculine and feminine gender. It is unclear at which point the markers were too infrequent or no longer formally distinctive enough to support the difference between the two genders.[4] More and more speakers lost the 'gender feeling' ("genusgevoel" E-ANS § 3·3·3·5), the knowledge which nouns are masculine and which are feminine. What remained was the distinction between the neuter definite article *dat* or *het* and its non-neuter counterpart *d(i)e*. The difference between neuter and non-neuter is marked suppletively rather than inflectionally, making it much less vulnerable to erosion.

3.1.3 Dealing with the loss: Word lists and dictionaries

The decrease in gender-distinguishing inflectional forms went hand in hand with an increasing uncertainty about which gender a noun belonged to. As early as 1584, the *Twe-spraack* voiced concerns about the speakers' (or rather writers') competence with regard to gender. In written language use, the author reported no "reghelmaat, nóch schickelyckheyd, in geslacht" ('rule nor appropriateness in gender', p. 69, author's translation).[5]

The educational elite reacted to this development by compiling word lists that marked nouns as masculine or feminine. These attempts more than anything proved that speakers were indeed no longer reliably aware of the distinction because they were not sufficiently exposed to unambiguous markers. The most influential word list was presented by van Hoogstraten in 1700. It contained some 1200 nouns and

---

[4] The same question could be asked for present-day Flemish, Brabantian and Limburgian. Given the extensive variation that seems to characterize the gender usage in these varieties, there are reasons to assume that their system is approaching the phase where markers no longer sufficiently support the three-gender system. Some Flemish data is discussed in 9.6 below.

[5] Originally "regel, maat" (edition Dibbets 1985: 225).

their gender. Gender affiliation was established on the basis of the usage by the honoured writers Pieter Corneliszoon Hooft (1581-1647) and Joost van den Vondel (1587-1679).[6] This list was reprinted and updated by various other scholars (see Rutten 2006 for the history of the word list). As time proceeded and the written and the spoken gender system drifted further apart, the matter moved from a grammatical to an orthographic issue and was mainly discussed in the context of spelling reforms. Notorious for its gender and case normativity is the orthography by De Vries en Te Winkel (1863) and their *Woordenlijst voor de spelling der Nederlandsche taal* (1866), the forerunner of what is now *Het Groene Boekje* 'the green booklet'. Issued by the *Nederlandse Taalunie* (Dutch Language Union) under the name of "Woordenlijst Nederlandse Taal", this book represents the official spelling dictionary, and is probably the most widely consulted source on grammatical gender.

The last of the inflectional suffixes on determiners and adjectives, the masculine *-n*, was declared optional by the orthography of Marchant (1947) and later by the *Groene Boekje* of 1954, "waarna er nooit meer iets van is vernomen" ("whereupon nothing more was heard of it", Van der Sijs 2004: 445, author's translation). The Dutch authorities finally embraced the system that had prevailed in the spoken standard language all along. Adjectives and articles now take the forms as in (3), both in the nominative and the oblique case (no case distinctions productively exist in present day Dutch outside the paradigms of the personal pronouns, except for a genitive-*s* which is restricted to proper names and terms of address such as *vader* 'father').[7]

(3) de   oud-e   man   een oud-e man
     DEF.C   old-C/N   man(C)   an   old-C man(C)

     de   jong-e   vrouw   een jong-e   vrouw
     DEF.C   young-C/N woman(C)   a   young-C woman(C)

     het   klein-e   kind   een klein   kind
     DEF.N   small-N   child(N)   a   small.N child(N)

As regards adnominal elements, Dutch no longer distinguishes masculine and feminine gender. From a linguistic point of view, this means that the two genders have merged into one. The group that combines the original masculines and

---

[6] Unfortunately, even the role models for gender usage proved to be unreliable. Vondel, born to Brabantian parents in German Cologne, differed quite often in his gender choice from Hooft, who was born and raised in Amsterdam. Besides, neither of the authors was fully consistent in his own work (see Kollewijn 1916 for comparison and critical discussion).

[7] Unproductive remnants of inflected forms are preserved in fixed expressions such as *de tand des tijds* 'time's tooth', *ter wereld* 'of/in(to) the world', *in koelen bloede* 'in cold blood'.

feminines is referred to as *common*, *uter*, *non-neuter* or *de*-gender (after the definite article), while the rest of the nouns are neuter or *het*-words (again, after the definite article).

Yet, dictionaries are hesitant to acknowledge the new gender. This has to do with the fact that the distinction between masculine and feminine is still formally marked on the personal pronouns. These pronouns, therefore, are expected to uphold the traditional split.

This view is supported by the linguistic behaviour of two groups of language users. The first consists of speakers of southern Dutch or Flemish dialects. In these dialects, many nouns are pronominalized as feminines, ranging from underived native words such as *bank* 'bench, bank' or *broek* '(pair of) trousers' to complex loanwords such as *organisatie* 'organization'. Feminine pronouns for the former group are a typical trait of southern dialects; northern Dutch speakers use masculine pronouns instead. The following reported dialogue illustrates the difference between north and south. A colleague jokingly exploits the difference between himself (A, from the north) and his wife (B, a southern dialect speaker) in the pronominalization of the noun *broek* 'trousers'. Note that the feminine pronoun (reduced form) is syncretic with the third person plural pronoun.

(4)   A: Heb je   mijn **broek**   gewassen?
            have you my   trousers.SG(C) washed
            'Did you wash my trousers?'

        B: Ja, ik heb **ze**   gewassen.
            yes I   have 3.F/3PL   washed
            'Yes, I washed it/them'

        A: Oh, heb je   meerdere **broeken**   gewassen?
            oh   have you several   trousers.PL washed
            'Oh so you've washed several pairs of trousers?'

        B: Nee, alleen de   jouwe.
            no   only   DEM.C yours
            'No, only yours'

        A: Dus je   hebt **'m**   gewassen?
            so   you have 3.M   washed
            'So you've washed them?'

        B: Ja, ik zeg toch, ik heb **ze**   gewassen.
            yes I   say AFF   I   have 3.F   washed
            'Yes, as I said, I washed them.'

The southern speakers' preference for feminine pronouns with *broek* 'trousers' is a direct consequence of the fact that southern varieties still mark masculine and

feminine gender on the determiners. In Southern Limburgian dialects, for example, *een brook* 'a.F pair of trousers(F)' contrasts with *ene rok* 'a.M skirt(M)' (Gaston Dorren, personal communication). For speakers of these varieties, the distinction between the two genders is still alive and governs pronoun usage.

The second group is formed by highly educated speakers of *Algemeen Beschaafd Nederlands* ('Common Civilized Dutch'), the official standard language. This group, many members of which can be characterized as "native speakers of written Dutch" (Jaap van Marle, personal communication), possesses active knowledge of the traditional gender system through schooling and extensive contact with literary and formal language. For many speakers of this group, it is normal to use feminine pronouns for certain groups of nouns, for example for abstract nouns ending in *-ing* (e.g. *lezing* 'reading'), *-age* (e.g. *lekkage* 'leakage') or *-heid* (e.g. *kindheid* 'childhood'). The full list of derivational patterns associated with particular genders is given in Table (5). In contrast to the southern Dutch dialects, underived native nouns such as *broek* '(pair of) trousers' and *bank* 'bench/bank' are pronominalized as masculines.

Table (5) Morphological patterns associated with masculine and feminine gender (*Woordenlijst Nederlandse Taal*, 1995)

Masculine	Feminine
nominalized verb stems	verb stems + suffix *-ing* or *-st*
nouns ending in *-aar, -aard, -er, -erd*	nouns ending in *-heid, -nis, -schap, -de, -te, -ij, -erij, -arij, -enij, -ernij, -ie, -tie, -logie, -sofie, -agogie, -iek, -ica, -theek, -teit, -iteit, -tuur, -suur, -ade[8], -ide, -ode, -ude, -age, -ine, -se, -sis, -xis, -tis*

Highly educated speakers come closest to what is traditionally the correct use of the pronominal genders. Yet, the *Algemene Nederlandse Spraakkunst* (*ANS*) - the standard reference grammar - admits that this usage is neither typical nor widespread in spoken language. It says "in de geschreven taal worden in het noorden een aantal, met name formeel gekenmerkte, substantieven ook wel als vrouwelijk behandeld" ('in the written language of the north, a number of nouns, most of them formally marked, are sometimes also treated as feminines', E-ANS § 3·3·3·4).[9] Thus, the

---

[8] Note that *chocolade* 'chocolate', which should be feminine according to its morphology, is listed as masculine in the *Groene Boekje* as well as in the *Van Dale* dictionary.

[9] The ANS also mentions the regional tendency to pronominalize mass nouns by means of a feminine pronoun. This usage is reported in Maljaars (1979: 107) as a trait of contemporary Dutch north of the great rivers. We will review this and other observations in Chapter 6.

feminine pronouns for derived nouns are actually a phenomenon of the written language, and marginal even then.

For most speakers outside of the two mentioned groups, and generally in colloquial speech, feminine pronouns for abstract nouns are unusual, and sometimes downright uninterpretable. In his witty and astute account, Jelle de Vries (2001) gives the following example (De Vries 2001: 101).

(6)   Met de **relatie** tussen Loes en Theo gaat 't mis,
        with DEF.C relationship(C) between Loes and Theo goes it wrong

        denk ik. **Ze** heeft kennelijk **haar** beste tijd gehad.
        think I 3.F has apparently POSS.F best time had

        'The relationship between Loes and Theo is going wrong, I think. It seems that it (lit.: she)'s past its (lit.: her) prime'

In the spoken language, the second sentence will be interpreted as indicating the decline of the female partner rather than, as intended, the relationship. By contrast, a common gender demonstrative conveys the intended meaning (De Vries 2001: 101):

(6)'   **Die** heeft **z'n** beste tijd gehad.
        DEM.C has POSS.M best time had
        'It's past its prime'

In the last decades, the two main authorities, the *Groene Boekje* and the main Dutch dictionary *Van Dale Groot woordenboek van de Nederlandse taal*, have tried to negotiate between normativity and description, written and spoken language and northern and southern varieties. In the attempt to integrate all the factors, the *Groene Boekje* currently distinguishes seven types of noun (*m*. stands for *mannelijk* 'masculine', *v*. for *vrouwelijk* 'feminine'):

- *de [m.]* (nouns that take the definite article *de* and are masculine, e.g. *vader* 'father', *ingang* 'entry', *rijkdom* 'wealth, richness')
- *de [v.]* (nouns that take the definite article *de* and are feminine, e.g. *moeder* 'mother', *universiteit* 'university', *gevangenis* 'prison')
- *de* (nouns that take the definite article *de* and are masculine or feminine, e.g. *bank* 'bench/bank')
- *de [m]_het* (nouns that take the definite article *de* and are masculine, or take the neuter definite article *het*, e.g. *aanrecht* 'kitchen unit')
- *de [v]_het* (nouns that take the definite article *de* and are feminine, or take the neuter definite article *het*, e.g. *idee* 'idea')
- *de* and *het* (nouns that take the definite article *de* or the neuter definite article *het* e.g. *matras* 'mattress')
- *het* (nouns that take the neuter definite article *het*, e.g. *boek* 'book').

A slightly different solution is chosen in the *Van Dale* dictionary, which distinguishes as many as eight different groups of nouns (again, *m.* stands for *mannelijk* 'masculine' and *v.* for *vrouwelijk* 'feminine', while *o.* is short for *onzijdig* 'neuter'):

- de; m (*de*-noun, masculine, e.g. *berg* 'mountain', *chocolade* 'chocolate')
- de; v (*de*-noun, feminine, e.g. *gevangenis* 'prison', *universiteit* 'university')
- de; m,v (*de*-noun, masculine or feminine, e.g. *mens* 'human', *persoon* 'person')
- de; v(m) (*de*-noun, feminine (or masculine), e.g. *broek* 'trousers', *bank* 'bench/bank')
- de, het; v(m) (*de*- or *het*-noun, feminine (or masculine), e.g. *matras* 'mattress')
- het, de; o en v (*het*- or *de*-noun, neuter or feminine, e.g. *idee* 'idea')
- het, de; o en m (*het*- or *de*-noun, neuter or masculine, e.g. *aanrecht* 'kitchen unit', *deksel* 'lid')
- het; o (*het*-noun, e.g. *kind* 'child', *boek* 'book')

Here, an additional distinction is made within the group of nouns that take *de* as the definite article and masculine or feminine pronouns. In reference to persons, pronoun choice varies according to natural gender. In reference to objects, preferences mirror differences between north and south, with the tendency of northern speakers to masculinize former feminines. This is expressed by the label "v(m)".

While the *Van Dale* dictionary has a lengthy help topic on grammatical gender, the current *Groene Boekje* does not provide any explanation for its decisions on the issue. Its criterion is how a word "wordt ervaren en gebruikt" ('is experienced and used', from the preface to the *Woordenlijst Nederlandse Taal* online). Unfortunately, this approach is partly a self-fulfilling prophecy. For the written language, the expectations are set by the normative rules of the past, which then re-enter the dictionaries as usage facts. This dilemma highlights the urgency of research on spoken language, where one can hope to get much closer to the speakers' grammatical reality.

For the linguist, there are both theoretical and empirical problems with the current official account. Theoretically, the two analyses can be criticized for treating *de*-nouns and *het*-nouns differently in what is used as evidence for the gender of a noun. Consider the *de*-nouns *vader* 'father', *rijkdom* 'wealth, richness' and *bank* 'bench/bank'. According to the dictionaries, *vader* is masculine because it denotes a male person. This is a semantic rule. Also, it takes masculine pronouns, which is a usage fact. *Rijkdom* 'wealth, richness', by contrast, is masculine because it ends in the suffix *-dom*. This is a morphological rule. *Bank* 'bench/bank', in turn, is masculine or feminine because it is pronominalized as masculine in the north and as feminine in the south. This is again a usage fact. So far, the choices are understandable considering that the Dutch gender assignment system is complex and cannot be explained on the basis of semantics or morphology alone. Yet, for *het-*

42

nouns only the determiners are considered. *Kind* 'child', is seen as a neuter noun because it takes neuter definite articles, although it often appears with masculine or feminine pronouns. Similarly, semantic rules are ignored. The *het*-noun *meisje* 'girl' is not considered a feminine noun, while the *de*-noun *tante* 'aunt' is, despite the fact that both refer to female persons (and take feminine pronouns by preference). The same holds for neuter nouns such as *boek* 'book', which often take masculine pronouns in colloquial speech. This usage is considered substandard, although it occurs frequently and systematically (see Chapters 5 to 8). Again, the dictionaries only count the determiner as evidence and ignore the usage facts.

Empirically, research on spontaneous speech, discussed in more detail in Chapters 5 to 8, shows that pronoun usage in northern Dutch does not quite correspond to the official account. The problematic group is the largest in the Dutch noun vocabulary: the nouns with inanimate reference. For the neuter nouns in this group, the ANS only states that they can appear with neuter pronouns. For the common gender nouns, the grammar says the following "Ten aanzien van *de*-woorden die geen personen of dieren aanduiden geldt het volgende. In de gesproken taal worden deze woorden in het noorden vrij algemeen als mannelijk behandeld" ('With regard to *de*-words that do not denote persons or animals the following holds. In the spoken language these nouns are generally treated as masculine in the north', E-ANS § 3·3·3·4). This impression, which has been voiced quite a few times in the relevant literature (see section 6.2), is not entirely confirmed by the corpus data. While many former feminines, e.g. *kast* 'wardrobe', *zon* 'sun', *broek* 'trousers', now take masculine pronouns in the standard language of the north, the masculine pronoun cannot be used for all *de*-nouns. The issue calls for systematic investigation, which is attempted in the present study. Moreover, it will be shown that it is not sufficient to discuss the former masculine and feminine nouns in order to understand the Dutch gender system of today. The pronominalization strategies for neuter nouns also play a significant part in the developments. Chapter 5 discusses in detail the usage patterns for each gender-marked pronoun.

3.1.4 Speaker awareness

The differences between written standard and speaker intuition have the consequence that language users are generally quite aware of their pronouns in writing. It is not unusual that the gender of a noun is actively debated.[10] The popular *language advice service* ("Taaladviesdienst") lists no less than forty different questions on grammatical gender, many of which regard pronominalization. The best reflector of this awareness are letters to the editor in newspapers. An example from the daily newspaper *De Volkskrant* criticizes the use of a common gender

---

[10] In fact, persons outside academia hearing about the present project have repeatedly mistaken the author, who is not a native speaker of Dutch, for an expert on 'the correct gender' and consulted her on the matter. Extreme opinions, heard in informal conversation, have even attributed speakers of Flemish or German with a superior knowledge of Dutch gender.

relative pronoun after the neuter noun *programma* 'program' and an instance of a neuter gender pronoun for the common gender noun *post* 'mail'.

(7) **Slordig**

Ik stoor me aan het groeiend aantal slordigheidsfouten in *de Volkskrant*. Een willekeurige greep uit de krant woensdag 5 mei:
> een **reparatieprogramma** *die* het mankement verhelpt
> dat **post** vooral zoek raakt omdat *het* in de verkeerde…

'Sloppy

I'm bothered by the increasing number of sloppy mistakes in the *Volkskrant*. An arbitrary selection from the newspaper Wednesday 5 May:
> a reparation program(N) that.C cures the problem
> that mail(C) is getting lost because it.N [is put] in the wrong...'

(*De Volkskrant* 10 May 2004)

Another example is a letter published in the rubric *Taalergernissen* ('language offences') of the periodical *Onze Taal* ('Our Language'), complaining about the use of common gender relative or demonstrative pronouns after neuter nouns. The nouns listed by the writer are *lichaamsdeel* 'body part', *vriendje* 'boyfriend.DIM', *verkiezingsbord* 'election poster' and *het Turkse leger* 'the Turkish army' (Onze Taal 1/2006: 35). Neither of the cases is predicted or explained by the rules of the ANS or the *Groene Boekje*.

The highest degree of public awareness is enjoyed by a phenomenon referred to as "*haar*-ziekte" ('*haar*-disease'), a particular usage of the feminine possessive (full form) *haar*. This usage will be discussed below in section 3.1.6. Like no other, it exemplifies the gap between spoken and written language: it is a phenomenon of written language that goes so strongly against the speakers' intuition that it has attracted general attention.

The situation is very different in spontaneous speech. Cases of self-correction or of explicit negotiation are rare. The corpus provides one amusing example:

(8) A: als je zo'n intelligente **kip** hebt heb je
       if you such_a intelligent.C chicken(C) have have you

       eigenlijk geen hok voor nodig. 'k bedoel dan blijft **ie** ook
       actually no cage for need I mean then stays 3.M also

       wel binnen de ruimtes die je **'m** geeft.
       AFF inside the spaces that you 3.M give

       'if you have such an intelligent chicken you don't really need a run for it. I mean then it'll stay in the room you give it'

> B: **Ze**.
> 3.F
> 'She.'

> A: Gewoon zeggen van nou blijf zitten dan doet **ie** dat wel.
> simply say of now stay sit then does 3.M that AFF
> 'Just say stay there and it will do that.'

> B: **Ze**.
> 3.F
> 'She.'

> A: Ja, een kip is een "ze".
> yes a chicken is a she
> 'Yes, a chicken is a "she".'
> (CGN session 513)

In informal contexts, Dutch speakers seem to use pronouns as unselfconsciously as in any other language. This is an interesting fact. Apparently, the spoken language has found its own solution for the problem of pronominalization. One of the main objectives of the present study is to identify this solution.

3.1.5 Avoidance

An interesting test case is written discourse of low formality, a genre where spoken and written preferences can be expected to clash. In such texts, there are certain nouns that speakers find difficult to pronominalize. These are generally common gender nouns, a fact that is not surprising given the mismatch situation sketched above. While for neuter gender nouns a neuter gender pronoun is always available, there is no common gender personal pronoun. Instead, speakers have a choice between a masculine and a feminine pronoun, but - as we have seen - problems with deciding which pronoun is the correct option. Speakers may then resort to avoidance strategies such as repeating the noun or choosing a pronoun that will relieve them of the task of deciding on a gender. The most useful elements for this latter option are demonstratives such as *deze* 'this (proximal)' or *die* 'that (distal)', which do not have separate forms for masculine and feminine gender. Thus, they fill the gap of the common gender pronoun, although they are still demonstratives and may thus introduce new problems.

Example (9), from an English children's book translated into Dutch, shows both avoidance strategies. The translation is clearly clumsy. First, there are two repetitions of the noun *melk* 'milk', then a demonstrative is used where the information structure of the sentence does not require it.

(9) Ze zette **een beetje melk** in de zon om zuur te laten
3.F set a little milk(C) in DEF.C sun(C) to in_order to let

45

worden,	deed	een	paar	gedroogde	dadels	in	de	rest	van
get	put	a	few	dried	dates	in	DEF.C	rest(C)	of

de	**melk**	om	**de**	**melk**	te	zoeten	en	zette	**die**
DEF.C	milk(C)	in_order	DEF.C	milk(C)	to	sweeten	and	put	DEM.C

in	de	schaduw	om	af	te	koelen.
in	DEF.C	shade(C)	in_order	off	to	cool

(Frances Temple "Dochter van de Bedoeïnen", 1997, Lemniscaat, translated by Anneke Koning-Corveleijn)

English original:
She set **some milk** in the sun to sour, and **some** she mixed with dried dates to sweeten **it** for drinking, then put **it** in the shade to cool.
(Frances Temple "The Beduins' Gazelle", 1998[1996], Harper Trophy Books)

Pronoun avoidance by noun repetition can assume astonishing proportions, witness example (10) (the example is translated rather than glossed because it does not contain any pronouns). Why *chocolade* 'chocolate' is such a problematic noun to pronominalize will be clear after Chapter 6.

(10) Verwarm **de chocolade** in 1-2 minuten in de magnetron op vol vermogen. Is **de chocolade** nog niet helemaal gesmolten? Even roeren en laten staan, dan smelt **de chocolade** vanzelf. Lukt het niet? Zet **de chocolade** dan nog (een paar keer) 20 seconden in de magnetron. Roer de kardemom door **de** gesmolten **chocolade**. Doe **de chocolade** in een schaaltje. Serveer de noten, stukken peer en druiven op een apart bord, om aan tafel in **de chocolade** te dippen.

'Heat the chocolate in the microwave on full power for 1-2 minutes. Hasn't the chocolate melted completely yet? Stir for a moment and leave, then the chocolate will melt by itself. Doesn't it work? Put the chocolate back (a few times) in the microwave for 20 seconds. Stir the cardamom through the melted chocolate. Pour the chocolate in a bowl. Serve the nuts, pieces of pear and grapes on a separate plate, to dip into the chocolate at table'
(supermarket magazine AllerHande 11/2006)

The use of the common gender demonstratives *deze* or *die* provides an alternative for awkward repetitious structures such as the above. Unfortunately, this evasion manoeuvre may jeopardize the semantics of the sentence. The reason is a difference in usage between personal pronouns and demonstratives in Dutch. In a sentence with two nouns, an ordinary personal pronoun can be interpreted as coreferent with either of the nouns, whereas the demonstrative normally picks out the second as antecedent (E-ANS § 5·6·3·3·1). (11) illustrates the difference.

(11) **Jan**_i gaat **Pieter**_j bellen, als **hij**_{i/j}/**die**_j thuis is
      Jan goes Pieter call if/when 3.M/DEM.C at_home is

   'Jan is going to call Pieter if/when he's home'

The preference for the second noun is even stronger when the demonstrative is the proximal *deze* (which has a more formal ring when used anaphorically and is rare in colloquial speech) rather than the distal *die* (E-ANS § 5·6·3·3·1).

If speakers use a demonstrative pronoun in order to avoid gender choices, the pronoun may end up referring to the wrong noun. The written-language example (12) is a case in point.

(12) De **mummie** zal eerst een **CT-scan** ondergaan voordat **deze**
      the mummy(C) will first a CT-scan(C) undergo before DEM.C

      tentoongesteld wordt in het Sakkara museum
      exhibited becomes in DEF.N Sakkara museum(N)

   'The mummy will first undergo a CT-scan before it is exhibited in the
   Sakkara museum.'
    (daily newspaper *Metro*, 04-05-05)

The demonstrative is clearly intended to refer to the mummy rather than the scan: yet, *deze* strictly speaking links back to *CT-scan*. The ambiguity is aided by the fact that both nouns belong to the common gender.

Examples such as this are evidence that pronominalization is a problem in writing. Despite interpretational hazards, the common gender demonstrative is preferred because it agrees with the noun while remaining noncommittal about the masculine or feminine that still exists in speakers' metalinguistic expectations but no longer in their mental grammar. The use of the common gender demonstrative as avoidance strategy is acknowledged in the standard reference grammar ANS (E-ANS § 3·3·3·6), and in some cases it is actually recommended. A schoolbook from 1982 (Klijnhout-Moerman and Feteris) advises to use the common gender proximal demonstrative *deze* 'this' in cases of uncertainty about the masculinity or femininity of a noun. Such advice reflects the helplessness that Dutch speakers or writers may feel about pronoun usage and thus confirms that there is a gender problem in present day Dutch.

3.1.6 Hypercorrection

If word-final /n/, as discussed in section 3.1.2, was indeed the last gender-specific case marker, this means that the last overt adnominal exponent of the difference between masculine and feminine - other than zero - appeared in the masculine paradigm. Thus, distinctly feminine elements in the NP had disappeared first of all. Also, there is the general tendency to pronominalize former feminines as masculine.

The issue will be revisited in Chapters 5 and 6, where the corpus data is presented and discussed. This masculinization has opened the doors to a movement in the opposite direction, viz. the promotion of the feminine pronoun to a marker of high style. For speakers from the north of the Dutch language area, it reflects the command of a distinction only marginally alive in the spoken language, but associated with educated writing. Thus, the feminine is the prime candidate for hypercorrect language use.

Indeed, in some registers the feminine is overused to such an extent that it has attracted public attention. The popular term "*haar*-ziekte" (*'haar*-disease') was coined to describe the proliferation of the full-form feminine possessive *haar* 'her' in unexpected places. This extensive use of *haar* can be attributed to language history on the one hand, on the other to hypercorrection and opportunism.

The *haar* in question is used mainly in journalistic and administrative writing, and it appears preferably with collectives such as *partij* 'party' or *volk* 'people'. Sometimes the antecedent is a true historical feminine, such as *regering* 'government', but in many cases it is not. *Haar* is occasionally used for historical masculines such as *staat* 'state', and in fact, it is quite often found in combination with neuter nouns. These are the cases against which the protest in the media is directed, and the relevant sentences are smiled at or frowned upon by native speakers of Dutch. Two examples are given under (13).

(13) a) dat **het   Nederlandse volk   haar**   soevereiniteit terug krijgt
         that DEF.N Dutch   people(N) POSS.F sovereignty   back gets
         'that the Dutch people regains its sovereignty'
         (http://presscenter.nl/Message/default.asp?NewsID=763&CatID=100)

    b) **Het   ijshockey**   in Nederland   probeert alles
       DEF.N ice_hockey(N) in the_Netherlands tries   everything

       om   **haar** imago te verbeteren
       in_order POSS.F image to improve

       'The ice hockey in the Netherlands is trying everything in
       order to improve its image'
       (*De Volkskrant* 05 January 2004)

The link to collectives probably has a historical source. In earlier Dutch, the form *haar* was used as a plural possessive form for all three genders (this place in the paradigm is now occupied by *hun* 'their') (Van der Sijs 2004: 448). This usage is still reported in the grammar of Ten Kate (Ten Kate, Noordegraaf and Van der Wal 2001 [1723] I: 471). In Middle Dutch, many collective nouns, especially derived words, had feminine gender. Since collectives are often pronominalized by plural pronouns, a reinterpretation from a plural to a feminine pronoun is a likely path of development.

However, there are reasons to assume that the present-day *haar* is more than the remnant of an old plural pronoun. Suspiciously, the modern usage of the feminine for referents other than female persons is much more frequent with the full form of the possessive, whereas the former plural pronoun also had reduced or clitic forms such as '*er* (Hogenhout-Mulder 1983: 31). Such reduced forms are also common for the ordinary contemporary feminine singular possessive referring to female persons (where they take the forms *d'r* or '*r*). By contrast, reduced feminine forms in collective contexts are rare. The overuse clearly concerns *haar*, not *d'r* or -*r*. Considering that the full form *haar* is typical for the written language, this can be an indication that this special use of the feminine possessive involves a certain level of formality and style. This opens up the way to hypercorrection, as well as to conscious use and exploitation.

Haar is not only special among the feminine possessives; there are similar restrictions on non-possessive feminine pronouns. The full form possessive appears in contexts were other feminine pronouns are mostly excluded. This is particularly striking in sentences containing *haar* next to another pronoun with which it shares its antecedent. It is not uncommon for the other pronoun to take a non-feminine gender despite the presence of the feminine gender possessive. (14a and b) are two examples. In (14a), a neuter antecedent is followed by *haar*, but the next personal pronoun is neuter again. Another remarkable case is (14b): the neuter personal pronoun and the feminine possessive stand adjacent to each other. Note that both cases are from published texts that no doubt have received a fair share of editing. (14c) illustrates the less common situation: not only the possessives, but also the (reduced) personal pronoun has feminine gender, while referentially belonging to a neuter noun.

(14) a) ***Kennislink*** is een **project** [...]. Sinds **haar** online gang in
      Kennislink is a   project(N)   since POSS.F online going in

      2002 is **het**   al   snel   gegroeid
      2002 is 3.N already grown quickly

      '*Kennislink* is a project [...]. Since its [lit.: her] going online in 2002 it has grown quickly'
      (job advertisement from Dutch National Graduate School of Linguistics)

    b) ... bevat   WARSTEINER Premium Pilsner   enkel natuurlijke
        contains WARSTEINER Premium Pilsner(N) only   natural

      ingrediënten en   het zuiverste bronwater.   Hierdoor verkrijgt
      ingredients   and 3.N purest   spring_water(N) by_this   gets

      het haar   karakteristieke, frisse smaak
      3.N POSS.F characteristic   fresh   taste

'... contains WARSTEINER Premium Pilsner only natural ingredients and the purest spring water. This is how it [lit.: she] gets its characteristic, fresh taste'
(advertisement supermarket magazine *Allerhande* 10/2006)

c) dat **het**   **FB**   het zou   betreuren
    that DEF.N faculty_administration(N) it   would   lament

    tov   **haar** promotoren, **haar** promovendicoördinator
    with_regard_to POSS.F supervisors POSS.F PhD_coordinator(C)

    en **haar** promovendi   dat **ze** op dit   moment
    and POSS.F PhD_students dat 3.F at DEM.N moment(N)

    wegens   geldgebrek   **haar** beleid   niet kan waarmaken?
    because_of money_lack(N) POSS.F policy(N) not can realize

    'that the FB [faculty administration] would be sorry that it [lit.: she] is unable to realize its (lit. her) policy with regard to its [lit.: her] supervising staff, its [lit.: her] PhD coordinator and its [lit.: her] PhD students because of lack of money?'
    (official communication, Vrije Universiteit Amsterdam)

While the feminine for common gender collectives may be the remnant of a historical usage pattern, the feminine pronoun for neuter nouns is in all likelihood a case of hypercorrection (Van der Sijs 2003: 269) Most recently, examples can be found that do not meet the criterion 'collective'. Such cases can be considered opportunistic attempts to exploit the high style associated with feminine pronouns. Two examples from advertising texts are given under (15).

(15) a) **Dit**   **tropische eiland** zal je   echt   verbazen met
       DEM.N tropical   island(N) will you really astonish with

       haar   hagelwitte stranden
       POSS.F hail_white beaches

       'This tropical island will really astonish you with its [lit.: her] snow white beaches'
       (advertisement in daily newspaper *Metro* 19-08-2004)

   b) Toch is dit   ontwerp persoonlijk, passend bij   het
       yet is DEM.N design(N) personal   fitting   with DEF.N

       huis   en **haar** omgeving
       house(N)   and POSS.F surrounding

'Yet this design is personal, matching the house and its [lit.: her] surroundings'
(http://www.siergras.nl/frameset/tuinontwerp.php)

Presumably, the writers of the advertisements wanted to boost their products by the use of the feminine *haar*. Another striking example, cited under (16), supports the hypothesis that such usage is indeed detached from any of the usual semantic properties associated with the feminine pronoun, in particular, the femininity of the referent. Clearly, "King Winter" is (metaphorically) male.

(16) Als **Koning Winter haar** intrede heeft gedaan...
      when king(C) winter(C) POSS.F entry has   done
      'when King Winter has made his [lit.: her] entry...'
      (http://www.bnbdeuitspanning.nl/omgeving.html)

The examples given above underline the fact that such cases have little to do with 'ordinary' gender agreement. After all, gender is usually thought of as a lexical feature, with agreement that cannot be manipulated at will for the sake of certain pragmatic effects. Moreover, we will see in Chapter 5 that this usage of *haar* for inanimate entities is not found in spontaneous speech. For these reasons, this particular phenomenon will be kept separate and will not be discussed in any more detail in the present book.

After this brief look at the history of the Dutch gender problem and its echoes in grammar writing and language usage, it is time to return to a more theoretical view on the issue. In the remainder of this chapter, we will have a brief look at the analytical challenges that the contemporary Dutch gender system poses for linguistic investigation.

3.2 Challenges for linguistic analysis

Generally speaking, the Dutch history of gender loss by syncretism is perfectly in line with well-known diachronic developments. There is cross-linguistic evidence that erosion of morphosyntactic markers affects some agreement targets before others: Priestly (1983), Marchese (1988) and Corbett (1991: 143) provide evidence in favour of a hierarchical order in which categories lose formal exponents. Erosion usually starts on the noun itself, before progressing to attributive elements, verbs and, last of all, to personal pronouns. Of course, the development does not necessarily carry through to all the positions in the hierarchy. However, the unaffected elements are predicted to be lower on the scale.

The Dutch facts confirm the typological pattern. Erosion has removed the gender markers on the noun and reduced those on the attributive elements, even if this is to different degrees. The indefinite article is the same for all genders. The definite article and the adjective can mark the distinction between common and neuter gender. The same two-gender pattern is found in the paradigm of the relative

pronouns. The personal pronouns, however, have remained unaffected by the erosion process: they mark three genders to this day.

(17) Dutch gender agreement and gender distinctions

Agreement target (type)	Attributive			Relative pronoun	Personal pronoun
Agreement target (category)	Indefinite article	Definite article	Adjective	Relative pronoun	Personal pronoun
Example forms	*een*	*de* [C] *het* [N]	*mooi-e* [C] *mooi* [N]	*die* [C] *dat* [N]	*hij* [M] *zij* [F] *het* [N]
Number of genders	0	2	2	2	3

It is cross-linguistically typical that pronouns preserve distinctions that are lost elsewhere. Well-known examples are the dual number in Old English, preserved longest in the personal pronouns of the first and second person singular, and the oblique case visible in forms such as *him*, *her* or *them* in Modern English. For gender, paradigmatic splits such as that of Dutch are surprisingly common amongst Germanic languages. Of the standard languages, more than half have lost one or all of their nominal genders, whereas all of them still possess (at least) three gender distinctions in the pronominal paradigms (see Chapter 9). Thus, the Dutch facts fit well within the Germanic context and beyond.

One of the immediate benefits of such typological knowledge is that it explains the role of the demonstrative pronouns. Demonstratives can be used as attributive modifiers in the NP. As such, they have participated in the deflection process. In contemporary Dutch, they pattern with the definite article and the relative pronouns in their gender marking and only distinguish common and neuter gender.

(18) Dutch demonstrative pronouns

	Proximal	Distal
Common	*deze*	*die*
Neuter	*dit*	*dat*

However, demonstrative pronouns can be used as anaphors, just like personal pronouns. In this function, they provide a way out of the mismatch problem: unlike normal personal pronouns, they have the same gender values as the nouns and can agree syntactically. Thus, they replace the missing common gender anaphor. As we have seen, this may cause speakers to use a demonstrative in many of the contexts where an ordinary pronoun would be expected.

In other respects, the Dutch situation is a hard nut to crack for linguistic analysis. One of the main challenges concerns the question of how many - and which -

genders the language has. Since gender marking has partly been reduced by erosion, different targets answer the question differently. This calls for a critical assessment of the theoretical tools that are applied to the data.

Analyses of gender systems generally build on the (often unquestioned) expectation that there is a homogeneous gender system with one gender per noun, which is visible on any relevant agreement target. Under this approach, all agreements are equally good gender evidence.

This works fine for languages such as Spanish or German where the gender is generally the same throughout the agreement targets.[11] Thus, changing the noun in the following German examples changes all agreeing elements: determiners, relative pronouns and personal pronouns. Such faithful formal covariance represents a high degree of canonicity in gender agreement (see 2.4 above).

(19) Wir fragten **den Mann**/ **die Frau**/ **das Kind,**
we asked DEF.M man(M) DEF.F woman(F) DEF.N child(N)

der/die/das mitgekommen war aber **er/sie/es** wusste nichts davon
REL.M/F/N come_along was but 3.M/F/N knew nothing of_it

'We asked the man/the woman/the child who had come along, but he/she/it knew nothing about it'

From this mini data set, it is easy to conclude that German has three genders.

The approach yields less straightforward results for Dutch, because changing the nouns in the parallel Dutch examples does not have an effect on all the targets in the same way.

(20) We vroegen **de man**/ **vrouw** / **het kind,**
we asked DEF.C man/woman (C) DEF.N child(N)

die/dat meegegaan was maar **hij/zij/het** wist er niets van
REL.M/F/N come_along was but 3.M/F/N knew there nothing of

With *man* and *vrouw*, only the pronouns change in dependency of the noun, whereas with *child*, all targets are affected. Thus, according to the articles and the relative pronouns, there are two genders – glossed as common and neuter – while according to the personal pronouns, there are three, the familiar masculine, feminine and neuter.

---

[11] Actually, heterogeneities do appear when having a closer look. In German, indefinite determiners only distinguish feminine and non-feminine, many adjectival forms are syncretic and possessives look the same for masculine and neuter.

In the European linguistic tradition, such difficulties are avoided by taking the definite article as the prime indicator for a noun's gender. Thus, *meisje* 'girl' is sorted among the neuters because it takes a neuter definite article, despite the fact that it often combines with feminine personal pronouns. The present study also embraces this method. Yet, we should dwell for a moment on the theoretical difficulties involved, and on alternative approaches.

Determining gender on the basis of the definite article means that attributive gender is regarded as better evidence than other agreements. In particular, it involves actively disregarding pronominal gender. Yet, Dutch has a tradition of calling common gender nouns ("*de*-words") feminines when they take a feminine pronoun (*de vrouw, zij* 'the woman, she') and masculines when they take a masculine pronoun (*de man, hij* 'the man, he'). Here, the pronominal marking is taken as indicative and the attributive gender is ignored. This is clearly problematic, especially if the methodological choices are unmotivated.

Yet, in the face of conflicting evidence we still want a solution for the whole system rather than for each agreement target in isolation. Paradigmatic mismatches do not sever the links between agreement targets completely. The noun *water* [N] 'water' always selects a neuter gender definite article, a neuter gender adjective, a neuter gender relative and a neuter gender personal pronoun, despite the differences in the paradigms, and we do want out analysis to capture this fact. Moreover, some instances of syncretism appear to weigh heavier than others. The lack of distinct forms that makes masculine and feminine gender indistinguishable in the noun phrase is pervasive in the system: it also appears in the demonstratives and the relative pronouns. By contrast, the syncretism between masculine and neuter is a speciality of the possessives. It does not compromise the distinction between neuter and non-neuter nouns, as they are richly supported elsewhere in the system. Neither does the complete syncretism of the indefinite articles present any danger to the distinguishability between common and neuter gender. By contrast, the missing formal distinction between masculine and feminine in most paradigms except for the personal and possessive pronouns does jeopardize this gender distinction, as we have seen: speakers can no longer tell masculine from feminine nouns. Thus, it makes sense to draw a line between pervasive and sporadic syncretism in the paradigms.

Another option would be to chart the syntagmatic agreement **combinations** for each noun. Of all the approaches, this one is most theoretically faithful to our contemporary notion of gender systems as systems of *agreement classes*. Agreement classes are determined on the basis of the agreement patterns triggered by a particular controller. Again, the approach can be illustrated with the help of German.

Most German nouns trigger the same agreement value on all relevant targets. These nouns can be easily sorted into the three familiar genders masculine, feminine and neuter. Yet, a number of nouns in German show variation in their agreement. In particular, the personal pronouns can agree with the conceptual properties of its referent as well as with the formal properties of the noun (this has been introduced

as *syntactic* vs. *semantic agreement* in section 2.2 above). The best known examples are neuter nouns denoting persons, which often pronominalize according to the sex of their referents. Those nouns could be set apart in their own genders, as in Table (21).

(21) Genders according to attested target combinations

Gender (example)	Syntactic agreement	Semantic agreement
masculine (*Mann* 'man')	masculine	-
feminine (*Frau* 'woman')	feminine	-
neuter (*Haus* 'house')	neuter	-
neuter feminine (*Mädchen* 'girl')	neuter	feminine
neuter masculine (*Oberhaupt* 'chief, head')	neuter	masculine

There are two problems with this approach. First, the last two genders only contain a handful of nouns, usually set aside as 'hybrids' (Corbett 1991: 183). Second, Table (21) is far from complete. In the realm of person reference, German has feminine gender nouns referring to male persons (e.g. *Memme* [F] 'sissy, wimp') which are pronominalized by a masculine pronoun. Conversely, masculine gender nouns for women (e.g. *Vamp* 'vamp, seductive woman', listed as masculine in Görlach 2001) preferably take feminine pronouns. Moreover, there are masculine and feminine nouns that can take pronouns of either gender (depending on the sex of the referent, e.g. *Waise* [F] 'orphan' or *Star* [M] '(film/music) star'). Finally, there are neuter nouns that can take either of the three genders (masculine and feminine again linked to the sex of the referent, as in *Opfer* [N] 'victim' or *Kind* [N] 'child'). This approach would attribute German with ten genders, which contradicts tradition as well as intuition because it puts more marginal cases on a par with what is frequent and typical.

For Dutch, such an approach is even more troublesome. As we will see in Chapter 5, **all** logically possible combinations of attributive and personal pronoun genders actually occur. This holds for reference to persons as well as for inanimate reference. The combinations can be schematized as in (22).

(22) Attested combinations of attributive and pronominal genders in Dutch

attributive   **personal pronoun**

                                    masculine
common   feminine
neuter   common
                                    neuter
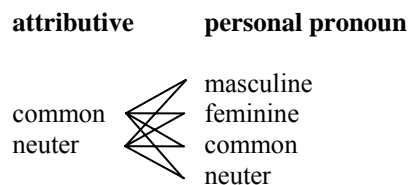
Not every noun can take all the possible agreements. However, it is easy to find examples for each combination. We will see many examples in the following chapters.

Worse, there are nouns that can take several different pronouns. In the extreme case, nouns can take any of the three pronominal genders plus the common gender demonstrative. Only a few nouns exhibit such pronominal promiscuity, and the true range is only achieved when written and spoken language are considered together, but examples are not hard to find. One is *muziek* [C] 'music'. It is a historical feminine, which may be reflected in written pronominalization. Elsewhere, we find common gender, masculine gender and neuter gender pronouns.

(23) a) Ik houd van   alle   **muziek**, als **die**   maar eerlijk is
          I   love   all   music   if   DEM.C   only honest is
          'I love all music as long as it's honest'
          (www.jazzmasters.nl/RADIO%20WEST.TWEE%20
          BLADZIJDEN%20BRECKER.doc)

    b) Ik maak **muziek** voor de   mensen die **het** mooi   vinden
          I   make music for the people who 3.N beautiful find

          'I make music for the people who like it'
          (www.marcoraaphorst.nl)

    c) Hij kent z'n **muziek**, hij kent **'m** al   gauw
          he knows his music(C) he knows 3.M already quickly

          uit   z'n hoofd
          from his head(N)

          'He know's his music, he knows it by heart quickly'
          (spontaneous speech, personal observation)

    d) de   klassieke moderne **muziek**, met **haar** complexe
          DEF.C classical modern   music   with POSS.F complex

          harmonieën
          harmonies

          'the modern classical music with its (lit.: her) complex harmonies'
          (www.radio4.nl/page/artikel/1631/)

Worse still, the combinations are even more diverse when relative and possessive pronouns are added. If we multiply out all the possibilities, the number of agreement classes explodes. This is the "maximalist problem" (Corbett 1991: 161 ff).

With the knowledge of the difficulties, we can return to the compromise solution outlined above. Attributive elements are not inherently superior to other agreement targets, but they are the most reliable evidence: nouns consistently take the same attributive gender, and this cannot be manipulated for the sake of semantic or pragmatic effects. Therefore, attributive agreement will be taken as indicative of the

lexical gender of a noun. All other manifestations of the same gender are regarded as instances of *syntactic agreement*. Thus, *boek* 'book' is a neuter noun because it takes neuter gender attributives, and in (24a), the pronoun agrees syntactically.

(24) a) Spannend **boek** uit? Ruil **het** gratis om
exciting.N book(N) out swap 3.N gratis around
'Finished reading an exciting book? Swap it for free.

Agreements with a different gender value, as in (24b) will be discussed under the heading *semantic agreement*.

    b) Spannend **boek** uit? Ruil **'m** gratis om
exciting.N book(N) out swap 3.M gratis around
'Finished reading an exciting book? Swap it for free.
(campaign on the website of the Dutch national railways, www.ns.nl)

Of course, this implies that the choice is indeed based on semantics. This is the topic of Chapters 5 and 6.

3.3 Conclusion

This chapter gives a brief history of the Dutch gender system. Until the Middle Ages, Dutch has three genders, but the language gradually lost its markers for masculine and feminine. This was part of a general process of deflection, i.e. erosion of inflectional suffixes. Modern Dutch now has the morphological means for only two genders in the noun phrase. The personal pronouns were spared in this proces. Yet, the increasing uncertainty about the gender affiliation of the former masculines and feminines affected their distribution. Despite educational efforts, speakers of the northern standard language lost the knowledge about masculine and feminine nouns, and ceased to use personal pronouns according to the historical genders. Heterogeneous behaviour in north and south, speech and writing caused trouble for lexicography and grammar writing, which also found it difficult to negotiate between description and prescription. The language users themselves, torn between intuition and norm, developed avoidance strategies and hypercorrect pronoun use.

Also for the investigating linguist, Dutch is a problematic language. The mismatching paradigms and the inconsistent agreements are a challenge to linguistic analysis. In this book, it is argued that determiners and adjectives are the most reliable indicators for the lexical gender of the noun. In this light, only common (*de*) and neuter (*het*) are syntactic genders in Dutch. The distribution of the four pronoun genders (masculine, feminine, neuter, plus common gender expressed on the demonstrative) in modern colloquial Dutch is the main blank spot on the map of the Dutch gender system and shall be investigated in the following. To this purpose, the results of the corpus study will be presented, analyzed and discussed. We will begin with a survey of methodological issues.

Chapter 4

Methodology

4.1 Data sources

The empirical heart of this book is a study of the Corpus Gesproken Nederlands (Corpus of Spoken Dutch, henceforward CGN), a 9 million word corpus of contemporary spoken Dutch and Flemish (for information and documentation, see http://lands.let.ru.nl/cgn/). Data was collected between 1998 and 2003 by a large number of collaborating institutions and the corpus was published in 2004 on 33 DVDs. For the present study, only the DVD was used that contained the corpus exploitation software COREX and the annotated transcriptions of the recorded sessions. The choice for the transcriptions rather than the original audio files was due to practical purposes: for collection and further use, the data was needed in written (digital) form.

For this study, a customized subcorpus was defined. In order to obtain the most natural data, the search was limited to spontaneous speech from face-to-face conversations (CGN text type *tta*). Other text types were regarded as less suitable because they contained planned, prepared and/or highly monitored discourse (broadcasts, simulated business negotiations, sermons, speeches and lectures, read speech).

Also, dialectal influence needed to be excluded, so the investigation was limited to the *Kerngebied* (core area), a term used in the corpus to cover the provinces Noord-Holland, Zuid-Holland and West Utrecht (see map below). Together, these provinces form the area known as the *Randstad*, the densely populated urban region in the west of the Netherlands. The varieties of the more rural east, northeast and south, as well as the regions of Brabant and Flanders (Belgium), were unsuitable for the present research because many of these dialects still have the traditional three-gender system and therefore lack the mismatch problem that is the focus of this project.[1] Moreover, nouns in regional varieties may have different genders than their cognates in the core area of Dutch. This last concern motivated the exclusion of data from Drenthe, Groningen and Friesland. In these northern- and easternmost areas of the Netherlands, the influence from Frisian and German, as well as from local dialects, was feared.

---

[1] Note, however, that Flemish and Brabantian exhibit similar tendencies as Northern Dutch in pronominalization. See Chapter 9 for details.

(1) Core area of the Corpus Gesproken Nederlands

(Map of the provinces of the Netherlands courtesy of Assembly of European Regions (AER.eu), used with permission)

With these regional and text type restrictions in mind, a subcorpus of around half a million words was constructed and searched. In the following, the term "corpus" or "CGN" is intended to refer to the subcorpus rather than to the CGN corpus as a whole.

In addition to the corpus data, examples from other sources were collected when encountered. They will be mentioned when they provide a particularly graphic illustration of a point. Such impressionistic data can be spoken language examples, but also cases of written text with literary, journalistic or advertising purposes: from books, newspapers, magazines, notice boards, on bottles, boxes and other

merchandise. In addition, examples from e-mail conversations and from the internet were (unsystematically) collected. Finally, observations from linguistic or popular writing on modern Dutch pronoun usage were considered. The examples from written language supplement the spoken data that forms the main body of evidence.

For any sort of counting or statistical calculation, only the systematically collected corpus data was used.

4.2 Corpus data: Search methodology

The setup of the corpus as a whole distinguishes three levels of structure. The first and largest is the *session*, which represents a single uninterrupted recording period. Sessions contain between 500 and 3000 words. The subcorpus used in this study is comprised of 257 sessions, giving around 500,000 words in total. For each example in this book, the session number is given in order to facilitate retrieval of the original utterance and its context. The second level of structure is the *turn*, meaning an uninterrupted contribution by a single speaker. This term is not used in the corpus setup, but it has practical value. "Turn" is also used as a measure of distance (see Chapter 8). The notion involves a certain degree of abstraction as turns from different speakers may overlap in time. Since the data was explored in transcribed form rather than listened to, the linear structure of the transcription was adopted as found. The third level of structure is the *sentence* or *utterance*. Utterances are defined by the corpus setup, and the decisions were not questioned.

The corpus was searched by reading through the conversations of the subcorpus and manually filtering out the relevant data. This method was necessitated by limitations in the corpus tagging and the exploitation software, the most serious being the absence of tagging for antecedent-anaphor relations. Furthermore, antecedent and anaphor are often widely separated, making it necessary to go beyond the concordance line of the search results and to access the conversation as a whole. In the light of these difficulties, reading through the subcorpus proved the best option. This type of search has the advantage of being - inaccuracies aside - exhaustive for the subcorpus in question.

Data was collected with two main research objectives in mind. The first aim was to establish the different sorts of referent for which a particular pronoun can be used. For this purpose, any pronoun was identified that differed in gender from its antecedent. Theoretically speaking, this part of the search holds the pronouns constant in order to establish the variety of nouns found as antecedents to these pronouns. There were roughly 900 of these gender-diverging pronouns in the subcorpus.

The second aim was to investigate under which conditions pronouns would switch, i.e. assume a different gender than their antecedent or earlier pronouns in the same agreement chain. For this investigation, all pronouns were relevant, those that diverged in gender from their antecedents and those that did not. Theoretically speaking, this means holding the nouns constant and collecting all agreeing

pronouns regardless of form or gender. This resulted in a database of 1629 pronouns with 550 unique antecedents. Examples from elsewhere, i.e. from other parts of the corpus and from own observations, furnish another 1300 cases. For each pronoun encountered, a record was entered into a Microsoft Access database.

4.3 Choice of pronouns

Dutch has a variety of pronominal elements, and not all of them are suitable for the purposes of the present research. Excluded were pronouns that do not agree in gender, such *iemand* 'somebody', *iets* 'something', *niemand* 'nobody', *niets* 'nothing', *diegene/datgene* 'the one who/which'. While some of them have links with gender (e.g. *diegene/datgene* contain the common and neuter gender demonstratives *die* [C] and *dat* [N]) and each of the listed pronouns triggers either common or neuter agreements on other elements, they are all distributed according to the semantic criterion of [person] versus [other]. This can be compared to English *what/who*, which also is not regarded as a gender distinction. For the same reason, the relative pronoun *wie* 'who' was excluded. By contrast, its neuter counterpart *wat* had to be considered since it often appears instead of the neuter gender relative pronoun *dat*. Some pronouns were too infrequent to figure in the data, for example the finite pronouns *ieder* 'each, any', *elk* 'each' in free pronominal use (they are vastly more common in attributive use).

Included in the search are personal pronouns, demonstrative pronouns in anaphoric use, relative pronouns, and possessive pronouns. Possessive pronouns in Dutch agree with the possessor rather than the possessum noun to which they are attributes. They are therefore treated on a par with syntactically distant agreement targets such as personal pronouns. Yet, possessives are special in that they can stand in the same clause as their antecedents, as in *Luise$_i$ asked her$_i$ brother*. The following paradigms give an overview over the relevant forms.

(2) Gender-agreeing pronouns in Dutch

Pronoun	**Case**	**Gender**		
		Masculine	Feminine	Neuter
Personal pronoun	NOM	*hij/ ie*	*zij/ ze*	*het[2]/ (e)t*
	OBL	*hem/ m*	*haar/ (d)r*	*het/ (e)t*
Possessive pronoun	NOM/OBL	*zijn/ z'n*	*haar/ (d)r*	*zijn/ z'n*
		Common		Neuter
Relative pronoun	NOM/OBL	*die*		*dat, wat*
Demonstrative	NOM/OBL	*deze*		*dit*
pronoun		*die*		*dat*

---

[2] While the full form *het* is the standard representation of the neuter definite article and pronoun in writing, its existence in the spoken language is questionable. The form *(e)t* is a more realistic transcription. Yet, in line with conventional orthography, *het* is used throughout the book.

Additional evidence comes from question words such as *welk(e)* 'which'; such cases were added to the database when encountered, but not included in calculations.

(3) a) of   gewoon **dat**   ene   **plankje**?   -
         or   simply DEM.N   one   plank.DIM(N)

         weet je   **welke** je   moet gebruiken?
         know you Q.C   YOU   must use

         'or simply that little plank? - do you know which you should use?'
         (CGN session 303)

     b) We   hadden wel zin   in   een **filmpie**,   maar
         we   had   PRT liking in   a   movie.DIM(N)   but

         wisten nog niet naar **welke** we   wilden gaan
         know   yet not   to   Q.C   we   want   go

         'We felt like seeing a movie, but we didn't know yet which one we wanted to go to'
         (http://www.silentpeople.nl/?m=200608

Two other types of evidence were considered in an off-the-record way. They are not pronominal and therefore only provide additional illustration. The first are full NP possessives (or *independent possessives*, in the terminology of the Dutch standard grammar ANS) which in Dutch are formed according to the template

[DEF POSS-*e*]

The definite article is marked for common or neuter gender and it is expected to agree with the possessum. Thus, the normal case is an utterance such as (4).

(4) Welk **koffiezetapparaat** nemen we, **het**   **mijne**?
     Q.N   coffee_machine(N)   take   we   DEF.N   POSS
     'Which coffee machine shall we use, mine?'

Instead, in real speech we often find examples where the agreement switches. (5) is a case in point.

(5) **'t**   **koffiezetapparaat**   -   ik   vergeet **de**   **mijne**
     DEF.N   coffee_machine(N)   I   forget   DEF.C   POSS

     vaak uit te   doen
     often off to   do

     'the coffee machine [..] - I often forget to switch mine off'
     (CGN session 252)

While such data is too interesting to ignore, there are two reasons to keep it separate. The first is, obviously, that the agreeing element is not a pronoun, but a definite article within a full NP. Second, the construction shares its derivational template with a certain kind of nominalization that is formally invariant and does not enter into agreement relations. For (5) above, the NP is disambiguated by the inanimacy of the referent, as *de mijne* as an independent nominal (i.e. not requiring an antecedent) can only refer to humans. In other cases, however, evidence can be inconclusive when it is not clear if the NP is an agreeing possessive NP or an invariant nominalization.

The same issue is relevant for the second, relatively large class of potential evidence. This class consists of elliptical noun phrases of the type illustrated in (6) below.

(6) ik   heb   nog   wel   een   **toestel**.   -   **een   goeie**?
     I   have yet   AFF   a   appliance(N)   a   good.C
     'I've still got a telephone somewhere. - A good one?'
     (CGN session 8192)

The elliptical noun phrase *een goeie* in (6) can be viewed as a gender switch. After all, the expected form of the adjective in the second sentence would have been the neuter *een goed*. Yet, *een goeie* (*een goede* in standard spelling) can also be a deadjectival nominalization. Such forms are homonymic to the common gender form of the adjective. They cannot provide any agreement evidence, as such nominalizations are formally invariant and thus do not agree.

Yet, there are a number of reasons why such constructions can be informative and should not be neglected. The first is that many cases clearly are elliptical NPs rather than nominalizations. This can be said with some certainty because nominalizations of the type [DET ADJ-*e*] are either of common gender and denote persons (*de blinde* 'the blind', *de lange* 'the tall one', *de onbekende* 'the stranger') or they are neuter gender nouns and refer to uncountable abstract entities (*het verwachte* 'the expected', *het vreemde* 'the strange thing, the foreign country', *het onbekende* 'the unknown'). Noun phrases that do not fit in either of the two classes must be elliptical. A very common case is *dezelfde/hetzelfde* 'the same' (the combination is so frequent that spelling treats it as a single word). Two other examples are (7) and (8). (8) has a plural antecedent, but the pronoun would be expected to agree with the gender of the singular base and have neuter gender. The treatment of such cases is discussed briefly towards the end of this section.

(7) een   **koffiezetapparaat**   is   dan   ook   niet   zo   duur.   -
     a   coffee_machine(N)   is   then also not   so   expensive

     nee   en   zeker   niet   **die**   **hele**   **simpele**
     no   and   certainly not   DEM. C   really.C   simple.C

     'a coffee machine is not so expensive. - No, and surely not such a simple one.'

(8) ik   heb   liever   vijf   kleine   **dingen**   gedaan
     I   have rather   five   small.PL thing.PL(N)   done

    dan   één   **grote**
    than one   big.C

    'I've rather done five small things than one big thing'

The adjectives in *de hele simpele* 'the very simple one' and *één grote* 'one big one' cannot be nominalizations; in the object or countable abstract meaning, they must have an NP antecedent.

Another reason why cases such as (7) and (8) above can count as evidence for a gender switch is that common gender adjectives do not occur with any arbitrary noun as antecedent. While they are not limited to common gender nouns, but are regularly used with neuters, not all neuter nouns qualify for such combinations. From the constructed examples under (9), only a) is possible. For b), the adjective must be neuter, as in c), or else the construction is avoided altogether.[3]

(9) a) Wat   voor   een   **pak**   trek   je   aan?   Een   **zwarte**
        what for   a   suit(N)   put   you   on   a   black.C

        of   een   g**rijze**?
        or   a   grey.C

        'What sort of suit will you be wearing? A black one or a grey one?

    b) *Wat   voor   **water**   wil   je?   **Warme**   of   **koude**?
        what   for   water(N)   want   you   warm.C   or   cold.C

        'What sort of water would you like? Warm or cold?

    c) Wat   voor   **water**   wil   je?   **Warm**   of   **koud**?
        what   for   water(N)   want   you   warm.N   or   cold.N

        'What sort of water would you like? Warm or cold?

The semantic patterns behind the combinability of neuter nouns with common gender adjectives in elliptical NPs are interesting and will be discussed in section 5.9.

In a similar off-the-record way, cases were added to the database in which the target is a singular pronoun of the type required for the present study, but in which the antecedent is plural. We have seen an example in (8) above, another is (10).

---

[3] Neuter deadjectival NPs such as *een goed* 'a good.N' are extremely rare.

(10)	Wilko	liep	naar	binnen	met	kleine	**kalfjes**		en	**deze**
	Wilko	went	to	inside	with	small	calf.DIM.PL(N)		and	DEM.C

	kwam	aanlopen	en	**die**	begon	aan	je	vingers	te	zuigen
	came	to_walk	and	DEM.C	started	on	your	fingers	to	suck

'Wilko went inside with a lot of small calves, and this one came and it started to suck on your fingers'
(CGN session 7836)

Such cases are clearly gender switches, and normally, a singular pronoun agrees with the gender of its antecedent, even when this antecedent appears in plural form. Section 1.2 above mentions the phenomenon of gender agreement under conditions of number mismatch. As we have no knowledge of the mental processing of such relations, in particular as we do not know if agreement conditions are looser when the antecedent is not explicitly present with the required features, records of this type were not included in statistical figures and calculations.

While all evidence mentioned in this section is relevant, only the personal, relative and possessive pronouns will be considered systematically.

4.4 Agreement conditions

Two agreement conditions were distinguished, the "match" and the "switch". Matching means that an agreement target shows the same gender value as its antecedent. An example for a match is (11).

(11)	hoe	heet	**dat**	**beest**	**dat**	in	Blijdorp	loopt?
	how	is_called	DEM.N	animal(N)	REL.N	in	Blijdorp	walks

'how's that animal called that walks around in Blijdorp (zoo)'
(CGN session 8175)

This sentence contains a neuter gender noun, identifiable as such by the preceding neuter gender attributive demonstrative, followed by a neuter gender relative pronoun.

All cases where agreement targets do not show the same gender as their antecedents are referred to as *switches*. Three random examples are given in (12).

(12)	a)	jeetje	wat	is	**dat**	**beest**	mager.	**hij**	heeft	gewoon
		EXCL	what	is	DEM.N	animal(N)	thin	3.M	has	simply

		een	dag	niet	gegeten
		a	day	not	eaten

'Gee, how thin that animal is. He just hasn't eaten for a day.'
(CGN session 8026)

b) een   slak   heeft een **huisje**   en   **die**   is   weer
    a   snail(C) has   a   house.DIM(N)   and   DEM.C   is   again

    gemaakt van kalk
    made   of   chalk(C)

    'a snail has a house and that in turn is made of chalk'
    (CGN session 9225)

c) ik   heb   nog   **kaas**   gevonden. dus   **dat**
    I   have yet   cheese(C) found   so   DEM.N

    ik   heb   **'t**   niet nodig
    I   have 3.N   not   need

    'I've found some cheese. So that, I dont need it.'
    (CGN session 7922)

Switches and matches can occur side by side within an agreement chain with several pronouns. The range of possible variation was investigated systematically and will be discussed in Chapter 8.

One particular type of antecedent was excluded from investigation. These were common gender nouns with human reference, such as *man* [C] 'man', *vrouw* [C] 'woman', *moeder* [C] 'mother', *leraar* [C] 'teacher'. Such nouns always take the pronouns that correspond to the gender of their referents or, alternatively, a common gender pronoun. The facts are unproblematic and well known. Also, variation was the key condition for the present study, and the only variation encountered in anaphoric reference to such nouns is that between ordinary pronoun and demonstrative, this being a different question than the gender switches which were investigated. Moreover, since the majority of the personal pronouns in the corpus refers to persons, such noun-pronoun combinations were also immensely frequent. Including them in the database was not practical and would not have provided any new insights.

4.5 Coding

The corpus data and the impressionistic data were archived in two databases. Per database record, a number of properties were coded for each antecedent-anaphor pair. The coding procedure for the corpus data is outlined as follows.

4.5.1 Number and text

Records contain the minimal piece of text containing antecedent and pronoun. The text is given in full, including all linguistic material between the two elements, excluding only those utterances that consist of nothing more than a non-speech sound such as *mmh*, *uhm* or *ggg* (the CGN code for laughter and other nonlinguistic utterances). Turn boundaries, i.e. changes of speaker, were marked with a hyphen. For the organization of turns, i.e. the linearization of the interaction, the choices of the corpus transcription were followed.

In the examples given in this book, utterances are slightly tidied and sometimes shortened. Half words and interjections are often unhelpful and disturb the orientation in the data. The database contains the full utterances and for counting, all words are considered.

If the conversation contains several pronouns for the same antecedent ("same" in the 'token' sense, not in the 'type' sense), the text cutout was chosen so as to include all relevant pronouns. Each pronoun was then assigned an individual record with a unique number.

4.5.2 Switch yes/no

The two agreement conditions, "switch" or "match", were coded binarily. All cases where the gender of the pronoun differed from the gender of the antecedent noun were counted as switches.

4.5.3 Antecedent and gender of noun

The antecedent of the pronoun was coded separately. If the pronoun referred exophorically to an implicit antecedent, the presumed antecedent was given in brackets. Naturally, this included some speculation, as it is impossible to know which entity, respectively which noun, the speaker had in mind. Exophoric pronouns were not included in any statistical figures.

Nouns were coded according to gender (common or neuter). Gender affiliation was established on the basis of the determiner the noun takes, as gender agreement is quite consistent within the noun phrase. Whenever doubts arose as to the gender of a noun in a particular speaker's grammar, the example was excluded. In particular, this was necessary for nouns with variable gender, such as *de/het matras* [C/N] 'the mattress'. An example such as the following (13) does not contain any clues as to which gender the speaker attributes to the noun: the sentence-initial demonstrative does not agree and the indefinite article does not vary in gender.

(13) dat   is een **luchtmatras**   maar **die**   wordt   bijgepompt
     DEM.N is a   air_mattress(C/N) but   DEM.C becomes refilled

'that is an air mattress but it's refilled'
(CGN session 254)

Thus, we do not know whether *die* in this sentence is a switch or a match, and the example has to be excluded.

4.5.4 Semantic class

The referents of noun and pronoun were sorted into semantic classes. The choice of classes was informed by typological knowledge about noun semantics and gender systems. The classes were the following:

- male person
- female person
- person
- male animal
- female animal
- animal
- bounded object/abstract
- specific mass
- unspecific mass/unbounded abstract
- collective
- event/place
- conjunct
- plural
- ambiguous
- uncertain

A short characterization of each class is helpful. The classes for MALE and FEMALE PERSONS and MALE and FEMALE ANIMALS include living beings whose natural gender is known to the speaker. Prototypical members of this group are specific persons and pets. Unknown natural gender is more common with animals than with humans, but it can occur in connection with unspecific referents, often with nouns such as *kind* 'child', *hoofd* 'head (e.g. of department)', *type* 'id.' or *slachtoffer* 'victim'. Persons with unknown gender were coded as PERSON, animals with unknown gender as ANIMAL. Unfortunately, the knowledge of the investigator does not always match that of the speaker. There are cases where the speaker is in all likelihood aware of the gender of the person or animal, as when speaking of a friend or the family dog, while the context fails to disclose this gender to the reader. In such cases, the PERSON or ANIMAL coding is used.

The BOUNDED OBJECTS/ABSTRACTS class includes referents that have clear conceptual boundaries, such as natural objects, artefacts, or body parts. Prototypical examples are a stone, a book or a hand. Boundedness can also be displayed by abstract entities. Bounded abstracts are, for example, a name, a piece of music, or an argument. Indeed, many nouns denoting bounded entities can have a concrete or an abstract reading. Especially in the realm of information technology, the boundaries

between concrete and abstract may be fluid. Is an e-mail on the screen concrete or abstract, and does it become less abstract when the same e-mail is printed? Computer memory can be bought as an object, but becomes an abstract property once installed. An article can be an abstract item on my CV or a concrete object on my desk. The question "Do you know this movie?" can refer to an abstract work of art, or to a concrete DVD. A letter, number or word can sit on a page or stick on a fridge door, but it is normally considered an abstract entity. The difficulty to tell concrete and abstract securely apart has motivated the choice of sorting bounded concretes and bounded abstracts together.

Opposite these bounded entities stands the group for UNSPECIFIC MASS/UNBOUNDED ABSTRACTS. Referents in this group are characterized as conceptually unbounded, "homogenous undifferentiated stuff without any certain shape or precise limits" (Koptjevskaja-Tamm 2004: 1067). The clearest examples are generic uses of mass nouns such as *honing* 'honey', *kleding* 'clothing' or *verf* 'paint'. Since mass nouns are usually open to recategorization as count terms (by a mechanism referred to as the *Universal Sorter* (Bunt 1985: 11) or the *Universal Packager* (cf. Jackendoff 1991: 24 on the unclear attribution of the term)), a formal restriction has been added to the semantic definition. In order to belong to this class, the nouns have to be used without determiner (except *geen* 'no', which combines with uncountable nouns). As in the bounded, countable class, this group contains concrete and abstract members. Among the uncountable abstracts we find nouns such as *informatie* 'information' or *gezondheid* 'health'. In general, such nouns are not often pronominalized and therefore only furnish a tiny number of examples.

In between the countables and the uncountables there is the class labelled SPECIFIC MASS. This class contains mass nouns that have acquired conceptual boundaries by denoting specific instantiations - i.e. types, sorts or portions - of substances. Relevant referents are a (sort of) wine, a (cup of) coffee, somebody's special cake (recipe). Nouns that look like mass nouns but are used with a determiner are also put into this group. Referents of such nouns have properties of masses on the one hand and of objects on the other and cannot simply be included in either of the two neighbouring classes. The three classes can be contrasted with the help of two constructed examples.

(14)   a)   (BOUNDED OBJECT)

Het	brood	ligt	op	tafel.
DEF.N	bread(N)	lies	on	table(C)

'the bread is on the table'

De	melk	staat	in	de	koelkast.
DEF.C	milk(C)	stands	in	DEF.C	fridge(C)

'The (bottle of) milk is in the fridge'

b) (SPECIFIC MASS)

Dit	brood	vind	ik	lekker
DEM.N	bread(N)	find	I	tasty

'This bread I like'

Voeg	de	melk	bij	het	deeg.
add	DEF.C	milk(C)	by	DEF.N	dough(N)

'Add the milk to the dough'

c) (UNSPECIFIC MASS)

Eten	we	nog	brood	erbij
eat	we	yet	bread(N)	with_it

'Are we having bread with this?'

We	hebben	geen	melk	meer.
we	have	no	milk(C)	anymore

'There isn't any milk left'

Nouns often have the propensity to appear in more than one of these classes, and quite a few cases are ambiguous. Ambiguities are frequently count-mass related. An example is (15).

(15)
dat	andere	**stokbrood**	smaakt	echt	heel	sterk	naar
DEM.N	other	baguette(N)	tastes	really	very	strongly	like

saté.	ik	vind	**deze**	echt	veel	lekkerder.	-
satay	I	find	DEM.C	really	much	tastier	

ja.	maar	hé	**'t**	vult	echt	onwijs.
yes	but	hey	3.N	fills	really	extremely

'Yes because that other baguette really tastes like satay. It's quite nice but I really like this one better. - Yes but hey, it's really extremely filling.'
(CGN session 468)

The noun is ambiguously a countable object or a specific mass, and there is no disambiguation throughout the conversation, although the context shifts the interpretation towards a true mass reading in the last sentence.

Another source of ambiguity comes from non-literal use of nouns, such as in metaphors or games. As the corpus contains several recordings during game-playing, there are quite a number of such cases.

(16)
ik	heb	een	**schaap**	-	heb	je	**die**	gepakt?
I	have	a	sheep(N)		have	you	DEM.C	taken

'I've got a sheep, don't you know that? - Have you taken it?'
(CGN session 422)

71

In this example, the sheep is a playing card and therefore ambiguously an object or an animal. In some cases, the difference is important. For reasons to be shown later, the pronoun choice in example (17) below is surprising and only explainable on the basis of the knowledge that *erts* 'ore' is in this case not a substance, but, again, a playing card.

(17) je   hebt   al   **erts**   gekregen. - **die**   had   ik   al
      you   have   already   ore(N)   got   DEM.C   had   I   already
      'you've already got ore. - I had that already'
      (CGN session 421)

When the class of the referent is ambiguous, it is marked as such.

Some nouns had to be marked "uncertain". These are cases where the referent was unclear, e.g. because there was more than one candidate, or because it could not comfortably be accommodated in any of the classes. The difference between cases marked as ambiguous and cases marked as uncertain is that the former fit more than one class, while the latter do not fit into any class at all. Neither uncertain nor ambiguous utterances were included in any statistical calculations.

The other classes COLLECTIVE, EVENT/PLACE, CONJUNCT and PLURAL were of less importance and were only added in order to accommodate interesting 'bycatch' to round off the picture of pronoun usage. Collectives here are institutions and other groups of humans conceptualized together as a body (there are also collectives of animals, such as flocks or herds, but the corpus yielded no examples). Relevant examples would be *bedrijf* 'company, firm', *gezin* 'family', *toneelgezelschap* 'theatre company' or *publiek* 'audience', but also *land* 'country', when the country was referred to as an political entity rather than a geographical place. EVENT/PLACE was used as a label for events such as *maansverduistering* 'lunar eclipse', *voorlichtingsdag* 'information day' or *communie* 'communion (day)' on the one hand and for places such as *tuin* 'garden', *bos* 'wood', *kerkhof* 'churchyard' on the other. Conjunct antecedents were *tomaat en mozzarella* 'tomato and mozzarella', *de ui en de knoflook* 'the onion and the garlic', *graan en erts* 'wheat and ore'. These needed to be considered separately because the individual nouns could differ in gender (and it could not be said which of the two served as antecedent) and because reference might be to the pair rather than to either of the two entities. In that case, anaphors such as *deze* 'this' or *die* 'that' would be plural and therefore genderless. Plural antecedent, finally, were *woorden* 'words', *aardappelen* 'potatoes', *boeken* 'books' and other nouns that appeared with singular pronouns in the corpus but are themselves plural.

Whenever the referent of the noun suggested a different class than the referent of the pronoun, the class of the pronominal referent was chosen.

4.5.5 Gender and sort of pronoun

The pronouns were coded according to sort, case, form, and gender. The choices were the following.

Sort: *personal - relative - possessive*
Case: *nominative - oblique* (only applies to personal pronouns, as the other pronouns do not have case morphology)
Form: *full - reduced - demonstrative* (full or reduced forms only apply to personal and possessive pronouns)
Gender: *masculine - feminine - neuter - common* (for personal pronouns)
             common - neuter (for relative pronouns)
             feminine - masculine/neuter (for possessive pronouns)

That demonstratives were counted as personal pronouns and merely coded as having 'demonstrative form' is due to practical reasons of the database setup rather than to theoretical considerations. The important point was that they could be separated from other pronouns for statistical purposes, which was guaranteed under the chosen layout.

4.5.6 Speaker and session

Each pronoun was linked to the person who produced it. Thus, the record for each pronoun gives the CGN code of the speaker, for example N01015. Furthermore, the session code was added to facilitate the retrieval of examples for readers wishing to go back to the original conversation. The software COREX also allows string searches, which in most cases provide the quickest access to the relevant place in the corpus. However, some examples in this book are slightly tidied, which is a problem for string searches. Then the session number is helpful.

4.5.7 Sequence and pattern

Often, the conversation contains several pronouns for one token antecedent noun. An example is (18).

(18) maar   het   **meisje**   **dat**   ik   begeleid   kan   dat
        but   DEF.N girl(N)   REL.N I   supervise   can   that

        helemaal   niet.   **die**   kan   **d'r**   eigen   veters
        at_all   not   DEM.C   can   POSS.F   own   shoelaces

        nog   niet   strikken   -   wat   raar.   hoe   oud   is   **ze**   dan?
        yet   not   tie   -   what   strange   how   old   is   DEF.F   then

        'but the girl that I supervise, she can't do that at all. She can't even tie her own shoelaces. - How strange. How old is she then?
        (CGN session 293)

In these types of cases, as many database records were created as there were pronouns in the text piece.

Under the headings of "sequence" and "pattern", the overall picture was captured. The "sequence" field gives all the record numbers belonging to the utterance (say 1010, 1011, 1012 and 1013). The "pattern" field relates to the switch/non-switch condition and gives the pattern of switched and matching pronouns for the antecedent in question. It was coded as *y* (yes, switch) or *n* (no, no switch). In the case of (18) above, there is a neuter gender antecedent and four pronouns, one matching (neuter), the others switched (common respectively feminine). Thus, the pattern is "nyyy". This coding procedure allowed identification of the range of variation in the switching behaviour of several pronouns for the same antecedent.

4.5.8 Syntactic distance and word-distance

One of the factors that are expected to influence switching behaviour is the distance between antecedent and controller. Choosing a suitable distance measure is a major challenge, theoretically as well as practically. Syntactic measures such as phrases are impractical due to the untidy syntax of the spoken language data. Counting clauses is a useful option for pronouns that are distant from their controllers, but it is too rough a measure for those that are closer. The method chosen was to code two distance measures, a rough *syntactic distance* and a fine-grained *word-distance*.

Syntactically, four levels of distance were distinguished:

- within clause (only possible for possessive pronouns)
- within sentence
- within turn
- beyond

Sentence and turn boundaries were taken as given by the corpus structure.

Left/right-dislocations such as (19a-b), which are fairly frequent in spoken discourse, were analysed as sentence- rather than clause-internal. This is in accordance with the commonly accepted syntactic analysis (as, e.g. in Lambrecht 2001) that treats the dislocated NP as an extra-clausal element (adjunct or argument). Thus, dislocated antecedents were coded as "within sentence".

(19) a) m'n **linkerbeen** **die** trilt heel erg
my left_leg(N) DEM.C shakes really badly
'my left leg, that's shaking really badly'
(CGN session 6897)

b) **die** heeft zoveel energie dat **beest**
DEM.C has so_much energy DEM.N animal(N)
'that one has got so much energy, that animal'
(CGN session 359)

74

Word distance was established by counting the words between the noun and the pronoun. Only full words were counted, excluding non-linguistic sounds. Distance could have the value 0 - meaning that antecedent and pronoun were adjacent - or any higher value. When the pronoun precedes its antecedent (cataphora), the word-distance was given as a negative value.

When there were several pronouns for a particular antecedent noun, each pronoun was coded for its own distance to the antecedent, although it is theoretically possible that pronouns influence or even agree with each other. If the noun was repeated, the last occurrence was used as starting point for the distance count.

The two distance measures combined provide evidence on the influence of word material between the antecedent and the pronoun, as well as that of larger structural boundaries such as between clauses, sentences and turns. Note that the two measures are principally independent: a word distance of zero (adjacency) can coincide with a turn change, which represents the maximal structural distance. While word-counting has its shortcomings in its blindness for structure, it has clear benefits on the side of practicality. Moreover, the main point was to compare the records to each other, thus establishing the relative rather than the absolute contribution of distance factors. To this purpose, consistency is the main criterion for the choice of method.

4.6 Excluded data

Both the nature of the investigated data and the particular research question put constraints on the material that could be used for the present research. This section briefly lists and motivates choices that have led to the exclusion of potentially relevant data.

First and most obvious, pronoun and antecedent had to share an agreement relation. For neuter pronouns in particular, this was not always the case. A common example would be copular constructions with predicative noun phrases such as the constructed utterance in (20).

(20) **Dit**   is   ons   **huis**.
      DEM.N   is   our.N   huis(N)
      'This is our house.'

The neuter pronoun *dit* in these types of constructions does not agree with the predicative noun, as changing the noun does not necessarily change the pronoun, witness (21).

(21) a) **Dit**   is   onze   **burgemeester/lerares**.
         DEM.N   is   our   mayor(C)/female_teacher(C)
         'This is our mayor/(female) teacher.'

b) **Dit**   zijn   onze   **kinderen**.
   DEM.N   are   our   child.PL
   'These are our children.'

The fact that neither a common gender noun nor a plural noun changes the pronoun shows that there is no agreement relation between pronoun and noun.

Also excluded were cases where the identity of the antecedent for a pronoun was ambiguous or unclear. A common source of ambiguities are noun phrases in predicative use, which often generate two possible antecedents, both being coreferent with the pronoun.

(22) De   **loodgieter**   is   zo   'n   klein   **mannetje**,
      DEF.C   plumber(C)   is   such   a   small   man.DIM.(N)

      die/dat   loopt   altijd...
      DEM.C/N   walks   always

      'The plumber is that sort of little man, he always walks...'

A common gender pronoun in (22) could be a switch (from the neuter noun *mannetje* 'little man', the closer of the two antecedent candidates), alternatively, it could reflect syntactic (non-switched) agreement with *loodgieter* 'plumber', the first noun. As there is no proof for either of the two alternatives, cases such as this were not considered.

Further, obvious complications are provided by idioms, collocations and other formulaic sequences that do not always respect the agreement relations normally required by the grammar. For example, (23) is an instantiation of an idiom pessimistically stating that something is not what it used to be.

(23) Mijn   **geheugen**   is   ook   niet   meer   wat   **het**   geweest   is.
      my   memory(N)   is   also   not   more   what   3.N   been   is
      'My memory is no longer what it used to be'

In this example, the pronoun *het* agrees with its antecedent *geheugen*, according to Dutch grammar. Yet, the agreement relation in these types of cases is often disrupted. Neuter pronouns seem to be the default choice in this template, and they are regularly used even with common gender antecedents. For nouns such as *toekomst* the different motivations of grammatical correctness and idiomaticity result in a bewildering variety of pronoun forms, as shown in (24). All forms (including the reduced forms of the personal pronouns) were encountered in an internet search. Interestingly, the syntactically agreeing common pronoun *die* is by far the least common (scoring only one hit in a search on www.google.com, as opposed to 37 hits for the masculine, 33 hits for the feminine, and 33 for the neuter).

(24) De **toekomst** is (ook) niet meer wat **het/'t/hij/'ie/zij/ze/die**
     DEF.C future(C) is (also) not more what 3N/N/M/M/F/F/C

    geweest is.
    been is

    'The future is no longer what it used to be'

Similar worries apply to idioms such as *op z'n beurt* 'in turn' (literally 'in his turn') or *op z'n kop* 'upside down' (literally 'on his head'). The possessive pronoun in these constructions often retains masculine gender even when the referent is female or plural. Idiomatic expressions have been avoided in this study by excluding the most obvious cases.

4.7 Switching nouns?

In all dealings with data, nouns and pronouns are taken at face value. This means that from the outset, the possibility has been ignored that if (say) a common gender pronoun follows a neuter gender noun, the speaker simply has had a different noun in mind. In example (25), this might, for example, be the common gender noun *kaart* 'card', as the wheat is a resource card in a game.

(25) ik wil **graan**. - nee **die** heb ik ook nodig
    I want wheat(N) no DEM.C have I also need
    'I'd like some wheat.' - 'No, I need that myself.'
    (CGN session 422)

Alternative nouns can be evoked for virtually all mismatching pronouns. Such reasoning is particularly suggestive when the antecedent noun is a diminutive with a common gender root. Common gender pronouns might then indicate that the speaker had the root noun in mind rather than the neuter gender diminutive they he/she just produced. Generally speaking, such objections cannot be refuted, as it is never possible to say with certainty on which noun a speaker bases his or her pronoun choice. Yet, there are empirical and theoretical ways of probing into the likelihood of such a scenario.

First, cases of gender-switched pronouns in Dutch occur with astonishingly high frequency. Chapter 8 gives some indications for just how often a pronoun will differ in gender from its antecedent noun. If all of these cases were slips of the mind, we would expect similar frequencies in other languages. Yet, Dutch is unusual in this respect.

More generally, if speakers were so prone to change their mind about the referents they are talking about or the nouns connected with them, it would be difficult to have systematic agreement at all. The tolerance for feature mismatches would be much higher, even in languages that do not have the paradigmatic problem sketched

in Chapter 3 above. The fact that Dutch is fairly unique in having a) a mismatch problem between the pronominal genders and the gender morphology elsewhere and b) massive gender-switching in the pronouns suggests that there is a correlation between the two. This makes the alternative claim - speakers have a different noun in mind - a less satisfying explanation since it does not make reference to the striking mismatch situation Dutch is saddled with.

Third and probably most importantly, having a different noun in mind should predict gender switching to be constrained solely by the availability of synonyms, semantic equivalents or at least close semantic neighbours that would lend themselves as substitutes for the antecedent noun. There should be no typical patterns in switches from one particular gender to another. Common-to-neuter and neuter-to-non-neuter switches should be equally common, and there should not be any general semantic rules predicting the likelihood and the direction of switching. However, Chapters 5 and 6 will show that the gender switches that do occur are everything but random. This points to a more systematic source than the accidental slip of the mind or the tongue.

Moreover, we will see that personal and relative pronouns differ in their propensity to diverge from the gender of their antecedents. This shows that switching is sensitive to target sort, rather than solely depending on the antecedent. The influence of the target has been observed for other agreement patterns, such as agreement with conjunct phrases (see Corbett 1991 and 2006 for discussion) and it is captured in the Agreement Hierarchy. Such 'alternative antecedents' have not been proposed for any of the phenomena related to hierarchy. It seems odd to invoke the argument only for Dutch and only in this particular mismatch situation.

As regards the diminutives, there is some support from psycholinguistic experiments that the gender of the base noun is activated along with the diminutive noun produced by the speaker, (e.g. Janssen and Caramazza 2003).[4] In order to control for the actual influence of the base noun gender on pronoun choice, diminutives with common gender bases were investigated separately in order to see if their liability for gender switching was higher than that of non-diminutives or diminutives with neuter bases. The results are presented and discussed in section 8.3.

---

[4] Note that the activation of the base noun gender has been tested in the context of determiner choice, with very simple, laboratory-style production tasks and under adjacency of noun and agreeing element. In the agreement phenomena investigated here, the influence of the gender of the root noun on the anaphoric pronoun is probably very limited. Two reasons come to mind. First, and most importantly, the wide separation of noun and pronoun in the discourse may obscure such subtle and possibly temporary influences as the gender of the base noun. Second, in natural, dialogical speech the speaker producing the antecedent noun and the speaker choosing an agreeing pronoun need not be the same person, and the activation patterns may be weaker in comprehension than in production.

In general, the methodology is chosen to produce a realistic pattern of the pronoun distribution in the corpus, and to provide a solid empirical basis for the theoretical questions ahead. Armed with the tools described in this chapter, we can now proceed to look at the data.

Chapter 5

Usage Data: Results of the Corpus Study

This chapter is concerned with 'wrong pronouns', i.e. with pronouns that diverge in gender from their antecedents. Such cases abound in Dutch spontaneous speech. The first aim of the corpus study was to chart the distribution of each of the pronouns in order to uncover the patterns behind the seemingly chaotic usage.

5.1 Preliminaries

Pronoun choice can be influenced by two factors: syntax and semantics. That is, a pronoun can be chosen on the basis of the lexical gender of its antecedent, or because of some semantic property of the referent.

Morphological and phonological properties of the noun should not play a direct role in pronoun choice. While they are known to influence the gender assignment to nouns, they are not expected to trigger a particular gender choice in pronominalization because syntax is expected to be morphology- and phonology-free (Zwicky 1969, Zwicky and Pullum 1986, Pullum and Zwicky 1988, Zwicky 1992). That means that agreement syntax is able to access the morphosyntactic features of the noun, such as its number or gender, but not its internal morphological or phonological structure.

Thus, wherever the syntactic gender of the noun is overridden in agreement, the reasons are expected to be semantic rather than morphological or phonological. An example illustrates this. Take the two Dutch nouns *meisje* 'girl' and *domoor* 'stupid person'. Both nouns have morphological and semantic gender assignment rules operating on them. In the case of *meisje*, the diminutive suffix *-je* advocates neuter gender, while the semantics - person reference - suggests common gender. The morphological rule wins out, *meisje* takes neuter gender attributives and is therefore considered to have neuter gender. Yet, in pronominalization, the choice may be revised in favour of semantics, and we often find feminine forms.

(1) Het   **meisje**   ... **zij**
     DEF.N girl.DIM.(N) ... 3.F

Thus, the syntactic gender of the noun can be overridden on semantic grounds in anaphoric agreement.

The reverse case, *domoor*, is an exocentric compound meaning 'fool' (literally 'stupid-ear'). It is headed by the neuter gender noun *oor* 'ear'. By virtue of the

81

Right-Hand Head Rule (Williams 1981), this should make *domoor* neuter, too. Yet, the fact that the compound denotes a person triumphs over the morphological rule, and the noun is assigned to common gender. Importantly, the losing morphological rule never gets a second chance in pronominalization. Since agreement rules cannot reach inside words, the internal structure of *domoor* is not accessible, so the noun is never found with neuter gender pronouns.

(2) De   **domoor**   ...   ***het**
     DEF.C   stupid_ear(C)   ...   3.N

Thus, those pronouns that diverge in gender from their antecedents are expected to do so for semantic reasons.

The present chapter is dedicated to the identification of the semantics behind gender switching in spoken Dutch. Switches can go in diverse directions, as illustrated in (3). For each of the two nominal genders, there are three possible switched pronominal genders.

(3) Directions of switching

noun gender	**pronoun gender**


```
                                    masculine
common                              common
neuter                              neuter
                                    feminine
```
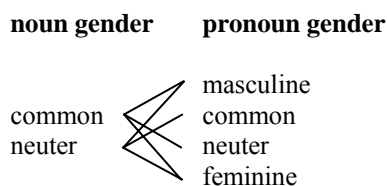
Each of the directions of switching were expected to have their own typical semantic pattern.

If no such systematic patterns can be established, the reasons for Dutch pronoun choice will have to be assumed to be syntactic after all. Apparent mismatches between nominal and pronominal gender will then be the result of some alternative syntactic mapping. For instance, it might be the case that speakers associate all common gender nouns with masculine gender in pronominalization. Switches from common to masculine would then be a matter of syntax rather than semantics. Of course, this leaves the question of how other switched pronouns are to be explained.

This chapter is structured as follows. The pronouns of the four genders – masculine, feminine, common and neuter – are each presented in an individual section. First we will look at the 'proper' personal pronouns, the masculine, the feminine and the neuter. For the masculine, the usage data will be complemented by a discussion of the claim that there is a syntax-based 'grammatical masculine' in Dutch. For the feminine pronouns, the particular use of the possessive pronoun *haar*, introduced in 3.2.5 above, will receive special attention. Last in this group, the neuter pronoun is discussed.

Section 5.4 will sketch the usage of the demonstrative pronouns *deze* (proximal, common gender), *die* (distal, common gender), *dit* (proximal, neuter gender) and *dat* (distal, neuter gender). After that, we will look at possessive pronouns, for which, unfortunately, there is little data. Finally, relative pronouns will be discussed. Both the demonstratives and the relatives are interesting evidence. Both pronoun types have the same gender values as the noun – common and neuter – yet, they switch according to the same patterns as the personal pronouns, albeit less often in the case of the relatives (the reasons for this are discussed in Chapter 8).

The chapter will conclude with a brief sketch of exophoric pronouns, i.e. of pronouns that do not have an overt antecedent. These cases are interesting because they may be the purest indicators for the particular referring potential of the various individual pronouns.

Some of the results of the corpus study replicate earlier findings or well-known facts, others are more newsworthy. Comparison of the results with the existing literature is deferred until Chapter 6. It should be borne in mind that the pronoun usage discussed in the present chapter concerns only ***switched*** pronouns, i.e. pronouns that do not have the same gender as their antecedent noun. This means that the question **how often** pronouns switch is not yet addressed (this is the topic of Chapter 8). Here, we are only concerned with the question: are there semantic patterns behind the switching?

5.1 Personal pronouns: Masculine

5.1.1. General: use of the masculine pronoun

The most straightforward use of the masculine pronoun, with the full forms *hij* (nominative) or *hem* (oblique) and the clitic forms *ie* (nominative) or *'m* (oblique) is for **male persons**. This is trivially the case with common gender nouns, but also holds true for neuters. Thus, masculine pronouns appear with nouns such as *broertje* 'little brother', *neefje* '(small) nephew, *jochie* 'small boy' or *mannetje* 'little man'. Two examples are given in (4).

(4) a) dat **jongetje** zegt **hij** vindt de   taal
         DEM.N boy.DIM(N) says 3.M finds DEF.C language(C)

         vindt **ie** moeilijk
         finds 3.M difficult

         'that little boy says that he finds the language difficult'
         (CGN session 9014)

    b) m'n **broertje**   en ik schelen twee jaar dus toen ik
        my   brother.DIM.(N) and I   differ   two   year so   when I

acht   was   was   **hij**   zes   en   toen   mocht   **hij**   ook
eight   was   was   3.M   six   and   then   was_allowed   3.M   also

om   acht   uur   naar   bed
at   eight   clock   to   bed

'My brother and I are two years apart. So when I was eight, he was six,
and he was also allowed to go to bed at eight.'
(CGN session 850)

This usage is particularly frequent with diminutives, as most nouns denoting persons
have common gender, so diminutives form the overwhelming majority of neuter
nouns with person reference. Yet, the usage is not restricted to diminutives.
Masculine pronouns also appear with non-diminutive neuter nouns such as *kind*
'child', *hoofd* 'head' (as in 'head of department') or *slachtoffer* 'victim'. In the case
of epicenes, which can refer to male or female persons alike, pronominalization
changes with the gender of the referent.

(5)   je   kan   't   **kind**   niet   een   ochtend   in   dat   huisje
      you   can   DEF.N   child(N)   not   a   morning   in   DEM.N   house.DIM(N)

laten   liggen.   en   je   kan   **'m**   ook   moeilijk   achter
let   lie   and   you   can   3.M   also   with_difficulty   back

in   een   zak   mee   op   je   rug   nemen
in   a   bag(C)   with   on   your   back(C)   take

'You can't leave the child in that house for a whole morning. But then you
also can't take him on your back in a bag'
(CGN session 358)

Other than for persons, masculine pronouns are used for **male animals** and for
animals whose natural gender is unknown or irrelevant to the speaker. This holds for
pets and farm animals, and intuitively also for lower animals, but there are no
examples in the corpus. The only reference to an animal not attached to a human
household is given in (6b). The animal in question is a wild bird outside the
speaker's window, and it can be assumed that she is unaware of its sex. (6a)
illustrates the more common situation: an animal of known sex, a male cat, is
referred to by a masculine pronoun, despite the fact that it is introduced as *dat beest*
'the animal', i.e. with a neuter gender noun.

(6)   a)   ja   op   gegeven   moment   was   dat   **beest**   helemaal   nat.
           yes   on   given   moment   was   DEM.N   animal(N)   all   wet

en   **hij**   zat   aan   't   infuus.
and   3.M   sat   at   the   drip

'Yes, at some point that animal was wet through and through. And he was on a drip.'
(CGN session 7837)

b)
omdat	die	**vogel**	zo	z'n	best	doet.[1]	dat	doet	**ie**	zeker
because	DEM.C	bird(C)	so	his	best	does	that	does	3.M	sure

hij	fluit	van	's	morgens	vroeg	tot	's
3.M	whistles	from	in_the	morning	early	until	in_the

avonds	laat
evening	late

'because that bird is so much doing its best. He's doing that for sure, because he's singing from morning till night'
(CGN session 518)

Perhaps more surprising, although noted in the grammars and a well-known fact to Dutch speakers, is the occasional use of masculine pronouns for female animals. In the example below, the pronoun refers to a female cat (the sex of the animal is deducible from the context which gives the cat's name and which contains a large number of feminine pronouns referring to the same individual).

(7)
arm	**beest**	-	ja	**hij**	kan	naar	buiten	he
poor.N	animal(N)	-	yes	3.M	can	to	outside	hey

dan	kan	**ie**	**z'n**	eigen	voedsel	gaan	vangen
than	can	3.M	POSS.M/N	own	food	ga	catch

'Poor animal. - Yes, he can go outside, can't he, and catch his own food.'
(CGN session 716)

The inconsistent choice for the masculine is further corroborated by example (8). Here, the neuter noun *beest* 'animal' is followed first by a feminine possessive and then by a masculine personal pronoun in the oblique case.

(8)
bedoel	als	zo	'n	**beest**	**'r**	rugje	doorgebeten	is
mean	if	such	an	animal(N)	POSS.F	back.DIM(N)	through_bitten	is

ja	dan	heeft	't	geen	zin	meer	om	**'m**	nog
yes	then	has	3.N	no	sense(C)	anymore	in_order	3.M	anymore

---

[1] The masculine possessive pronoun *z'n* is also potentially an agreeing element, but as it is part of the idiomatic expression *z'n best doen* 'do his best', it is not counted as evidence (see 4.5.6).

te   laten   rondlopen.
to   let   around_walk

'I mean if the back of the animal is bitten through then it's useless to let it walk around any more'
(CGN session 6741)

The usage of masculine pronouns for female animals such as *koe* 'cow' or *kip* 'chicken' is well documented, it is also a common feature of dialects (Geerts 1995b, Wahrig-Burfeind 1989). The corpus contains an example where a (hypothetical) chicken is referred to by masculine pronouns.

(9)   als   je   zo'n   intelligente **kip**   hebt   heb   je   eigenlijk
       if   you   such_a intelligent.C chicken(C)   have   have   you   actually

       geen   hok   voor   nodig. 'k   bedoel   dan   blijft **ie**   ook   wel   binnen
       no   cage   for   need   I   mean   then   stays 3.M   also   AFF   inside

       de   ruimtes   die   je   **'m**   geeft.
       the   rooms   that   you   3.M   give

       'if you have such an intelligent chicken you don't really need a run for it.
       I mean then it'll stay in the place you give it'
       (CGN session 513)

Leaving the animate domain, the masculine pronoun appears with a large range of inanimate referents. In the overwhelming majority of cases, the referent is an **object**.

(10) a)   dat   **masker**   dat   je   ouders   hebben   gekocht   vind   ik
           DEM.N mask(N)   REL.N your parents   have   bought   find   I

           niet   zo   nou   ja...   **hij**   is   wel   leuk   maar
           not   so   now   well   3.M   is   quite nice   but

           'that mask that your parents have bought, I don't really think its so... it's quite ok but...'
           (CGN session 469)

     b)   hier   heb   je   mijn   **apparaat,** ik   wil   **'m**   opwaarderen
           here   have   you   my   device(N)   I   want   3.M   top_up

           'Here you've got my telephone, I want to top it up'
           (CGN session 8132)

In other cases, the referent is an abstract entity.

(11) a) ja   ik weet een heel leuk **verhaal** over m'n
yes I   know a   really nice.N story(N) about my

broertje.   - je hebt **'m** al   'ns verteld
brother.DIM(N)   you have 3.M already once told

'Yes I know a really nice story about my brother. - Yes, you've told it before'
(CGN session 685)

b) lees **'m** nog 'ns voor 't   **argument**.
read 3.M again once out DEF.N argument(N)
'read it out again, the argument'
(CGN session 9227)

c) ben benieuwd of 'k **tentamen** gehaald heb. - meestal als
am curious   if I exam(N)   passed have   mostly when

je d'r   positief over bent dan haal je **'m** niet.
you there positive about are   then pass you 3.M not

I'm curious if I've passed the exam. - most times when you're positive about it then you don't pass'
(CGN session 662)

For yet other referents, it is not so clear if they should better be characterized as concrete or abstract (recall the discussion in section 4.5.4 above).

(12) a) ken je 't   **liedje**   niet? **hij** is echt   leuk.
know you DEF.N song.DIM.(N) not   3.M   is really nice

'k heb **'m** één keer gehoord. ik vond **'m** echt   leuk
I   have 3.M one time listened   I   found 3.M really nice

'Don't you know that song? I heard it once and I thought it's really great.'
(CGN session 686)

b) zodra   't   **mailtje**   ze zou   zien dan delete je
as_soon_as DEF.N e-mail.DIM(N) she would see   then delete you

'm zo snel   mogelijk
3.M as quickly possible

'as soon as she sees the email [...] then you should delete it as soon as possible'
(CGN session 504)

c) ik krijg betaald niet per **artikel** maar per aantal regels
I get paid not per article(N) but per number lines(N)

Dus ik zou **'m** zo af kunnen sluiten
so I could 3.M.OBL so off can close

'I'm not paid for each article, but for the number of lines. So I
could just finish it like this'
(CGN session 151)

The common characteristic of all these referents is that they are bounded, discrete
entities, whether concrete or abstract. In the following, such referents are discussed
under the heading *bounded object/abstract*.

The three classes - male persons, all animals and bounded objects/abstracts -
together constitute the semantic field covered by the masculine pronoun. That is, if a
neuter noun is followed by a masculine anaphor, the referent is expected to fit into
one of these classes.

5.1.2 Specific restrictions and preferences

So far, all masculine pronouns have been considered together. However, the usage
facts vary among the particular pronouns involved. Specifically, restrictions apply to
the use of full-form, i.e. non-clitic pronouns, and there is sensitivity to case.

While the nominative full form, *hij*, appears across the entire range of referents, the
full form oblique pronoun *hem* is very rare for inanimate entities. There are only 2
examples in my database, one of which is exophoric, i.e. without overt antecedent.

(13) a) hij had een **werkvergunning**. nou had ie geld betaald
he had a working_permit(C) now had 3.M money paid

en nou had ie **hem** wel.
and now had 3.M 3.M.OBL AFF

'He has a working permit. [...] Now he had paid money and now
he did have it.'
(CGN session 313)

b) die moet je voor mij wel even bijbestellen want
DEM.C must you for me order because I want

ik wil **hem** wel per se
I want 3.M.OBL AFF per se

'You must order a copy of that (a photo) for me, because I really want it'
(CGN session 388)

Moreover, the exactness of the transcription can be questioned, as the difference between clitic /əm/ or /ɛm/ and free word /hɛm/ is subtle. While inconsistencies in transcription could also hide a higher number of occurrences of the oblique full form pronoun, expectations are that this is not so. The standard grammar ANS claims that "the unstressed full form *hij* is in principle the only full form that can easily be used to refer to non-persons" (E-ANS § 3·3·3·4). This is clearest in clause-initial position where the reduced '*m* is not available because it is an enclitic and needs a host to attach to, and where /hɛm/ is less easily mistaken for /əm/ or /ɛm/. Here, the oblique pronoun is always interpreted as referring to a person (see constructed example (14a)). (14b) elicits no such bias.

(14) a) **Hem**   heb ik gisteren   nog   gezien.
          3.M.OBL have I   yesterday only   seen
          'Him, I've seen only yesterday'

    b) Ik heb **'m**   gisteren   nog   gezien.
          I   have 3.M.OBL yesterday only   seen
          'I've seen him/it only yesterday'

However, the oblique pronoun only appears sentence-initially when it is stressed, e.g. in contrastive contexts. This may explain the bias, as reference to inanimates seems to be restricted to unstressed full forms, as the standard grammar says (E-ANS § 3·3·3·4, see quote above). Examples such as (15a) (capitals indicate stress) are automatically taken as referring to an inanimate entity. By contrast, a demonstrative pronoun under the same syntactic and prosodic conditions (as in (15b)) can refer to inanimate or animate entities alike.

(15) a) Als **HIJ**   er   nog is...
          if   3.M.OBL there still is

    b) Als **DIE**   er   nog is...
          if   DEM.C there still is

          'If he/that's still there...'

Thus, the restrictions on the use of the full form oblique pronoun can be attributed to its stress rather than its case. Contrastively stressed personal pronouns are only used to refer to persons (and occasionally to higher animals), not to inanimate entities. Unfortunately, stress information could not be taken into account in the corpus study, so the point awaits further investigation.

5.1.3 Grammatical vs. semantic masculine

This section so far concentrates on neuter antecedents. With a neuter antecedent, a masculine pronoun is unambiguously the result of a switch, i.e. of an override of the syntactic gender of the noun. The question arises how to treat masculine pronouns

occurring after **common** gender nouns. After all, at least half of the common gender nouns are historically masculines, so masculine pronouns occurring with common gender nouns could be remnants of this system. Alternatively, the masculine could generally be mapped onto common gender, either because common gender nouns are in fact masculines, or because of some agreement rule, potentially a default rule.

A general correspondence between common nominal gender and masculine pronominal gender is a popular scenario in Dutch literature on the subject (Kruisinga 1924, Dekeyser 1980, Geeraerts 1992, Geerts 1966, 1995a and c, Van der Sijs 2004). The hypothesis is attractive. As we will see in the next section, standard Dutch has lost its ability to refer to inanimate entities by means of a feminine pronoun. To many minds, this suggests that the feminine's loss was the masculine's gain, with the result that practically all non-neuter nouns, except those with female referents, have adopted masculine gender. Besides, masculine gender pronouns occurring after common gender nouns are such a frequent phenomenon that a simple correspondence rule would seem to cover the data in the most parsimonious way.

The case, of course, must be decided empirically: if the data shows semantic restrictions on the use of the masculine pronoun, we can assume that it is semantics rather than syntax that governs the distribution of masculine gender agreement.

Unfortunately, the answer given by the data is less clear than might have been hoped for. While neuter nouns take masculine pronouns only under easily defineable semantic conditions - the referent needs to be animate or a bounded object or abstract - common gender nouns take masculine pronouns in a number of cases that go beyond these conditions. Thus, we find 22 examples where the masculine appears in a semantic field where it would not, were it an anaphor to a neuter noun. This is the field of mass nouns. Two examples are example is given in (16).

(16) a) wil je de **soep** effe in de koelkast zetten
         want you DEF.C soup(C) PRT in the fridge   put

         dat ik **'m** morgen effe opeet of zo.
         that I   3.M   tomorrow PRT eat_up or so

         'Can you put the soup in the fridge, for me to eat it tomorrow or so'
         (CGN session 990)

    b)   dat die **kaas** weer zo in de koelkast ligt.
         that DEM.C cheese(C) again so in the fridge   lies

         dat **ie** helemaal uitdroogd is dadelijk
         that 3.M   all   dried_out is soon

         'that the cheese is in the fridge again like that. That it's all dried out soon'
         (CGN session 498)

Yet, there is an important restriction on these nouns. They do denote substances - liquids, foodstuffs, materials - but in all cases these are specific instances of substances. In (16a), the soup is a quantity of soup in a container (which can be put in the fridge). The noun is accompanied by a determiner, which also indicates that this is not a prototypical mass reading. In (16b) the demonstrative *die* 'this' invokes a piece or sort of cheese rather than cheese in general. In yet other cases, we find the possessive *mijn* 'my', which also singles out a particular portion or sort.

It is important to consider the data carefully. (17) below at first sight looks like an example of a generic mass noun with a masculine anaphor, but there are two difficulties. On the level of form, the pronoun is ambiguous: a masculine clitic attached to a word ending in /t/ is phonetically identical to a common gender demonstrative, an element with much looser semantic restrictions. Semantically, the rest of the conversation makes clear that it is a specific sort of coffee that made the speaker go off coffee in general. Thus, the pronoun has a specific rather than a generic meaning after all.

(17) ik   ging   steeds minder **koffie**   drinken. en   dan   denk ik
     I   went   always less   coffee(C) drink   and   then   think I

     waarom is dat toch maar dat is omdat   **ie**   niet lekker is
     why   is that PRT   but   that is because 3.M not   tasty   is

     'I was drinking less and less coffee. And then I think, why is that, but that's because this coffee is no good'
     (CGN session 626)

Thus, the masculine pronoun occurs with referents of the class labelled "specific mass" rather than the class of 'true' or "unspecific" masses (remember section 4.5.4). Unbounded, uncountable abstracts are not found with masculine pronouns either.

Two generalizations emerge from the data. First, the semantic field for masculine pronouns following common gender nouns is slightly larger than that for masculine pronouns following neuter gender nouns. This suggests that speakers regard a switch from common to masculine as a smaller step than a switch from neuter to masculine. Yet, the distribution of the masculine pronoun is still defineable in semantic terms, and it obeys semantic restrictions: generic mass nouns are not pronominalized as masculine, no matter what their gender. Thus, the situation cannot be analyzed as a simple syntactic correspondence between common nominal gender and masculine pronominal gender. Besides, masculine pronouns with mass reference are rare. The 22 cases in the database contrast with 477 that refer to objects. Thus, the data suggests that masculine pronouns occurring after common gender nouns can be seen as instances of semantic rather than syntactic agreement. Speakers seem to make the switch from common to masculine more easily than from neuter to masculine, but their behaviour is better explained in semantic than in syntactic terms.

5.1.4 Distribution

The masculine pronouns in the database are distributed over the mentioned semantic classes in the following way (Table (18)). The distribution for neuter and common gender antecedents is given separately. Common gender nouns referring to persons are not considered.

(18) Masculine pronouns across semantic classes

Class	Male person	Male animal	Female person/ animal	Animal (general)	Bounded object/ abstract	Spec. mass	Unspec. mass/ abstract
Noun [N]	39	4	-	6	130	-	-
Noun [C]	×	5	3	27	347	22	-

Moreover, there are six cases in which the antecedent is a neuter noun with plural number; all of them denote bounded objects. Two instances, one with the proper name *Windows 98* as antecedent and one speaking of a school, did not fit any of the semantic groups. Finally, there is one instance where the antecedent denotes a female person. The example is given in (19).

(19) nou   hoorde ik   van   een   **meisje** dat   **ie**   gaat   trouwen
      now   hear   I   from a   girl(N)   that 3.M   goes   marry
      'Now I heard from a girl [...], that she's going to get married'
      (CGN session 631)

I attribute this case to a transcription error. The nominative masculine pronoun is phonetically identical to the common gender demonstrative when the preceding word ends in a /t/. Probably, the correct transcription would be *dat die*. In the standard language, the masculine pronoun is not used to refer to female persons.

The corpus data shows that the masculine pronoun can be used for nearly any sort of referent. Yet, its distribution is not random: we do not find masculine pronouns with unspecific masses and abstracts. This suggests that the masculine switches do obey semantic restrictions. Analyzing the data in semantic terms is therefore profitable.

5.2 Feminine

5.2.1 Use of the feminine pronoun

The feminine pronouns appear in the nominative forms *zij* and *ze* 'she' and in the oblique forms *haar* 'her' with the reduced versions *'r* and *d'r*. Their usage is very straightforward: in spontaneous speech, feminine pronouns always refer to female persons or - more rarely - animals. The most common case is feminine agreement with *meisje* 'girl', a diminutive of low transparency. A monomorphemic neuter

antecedent would be *wijf* 'woman (pejorative)', but the corpus does not contain any pronominalizations of this noun.

(20) a) een **meisje** met 'tzelfde badpak. en **ze** is ook blond.
a girl(N) with the_same.N swimsuit(N) and 3.F is also blonde
'a girl in the same swimsuit. And she's a blonde, too.'
(CGN session 383)

b) mijn **zusje** is dit jaar geslaagd dus 'k ging effe
my sister.DIM(N) is this year passed so I went PRT

met **'r** mee **d'r** cijferlijst halen
with 3.F with POSS.F mark_list get

'My sister has passed her exams this year, so I went with her to get her marks list.'
(CGN session 619)

The pronouns in both sentences are probably hardly perceived as switches. For (20b), syntactically agreeing, i.e. neuter pronouns would be extremely odd. For the possessive, the rejection of the neuter is probably aided by the fact that the neuter possessive is homophonous with the masculine (both have the form *zijn*). When pronouns acquire stronger semantic associations, it is likely that a neuter pronoun homophonous with a masculine pronoun feels increasingly less suitable for a female referent.

With regard to animals, the corpus yields only very few relevant cases (none of them are part of the subcorpus, so the search was extended beyond the 500,000 words that constituted the main sample). One example is (21), where the antecedent is the neuter noun *beest* 'animal' and the referent is a female cat.

(21) arm **beest** in elk geval als **ze** naar buiten gaat dan
poor animal(N) in any case if 3.F to outside goes then

komt **ze** d'r niet meer in
comes 3.F there no more in

'poor thing, in any case, when she's going outside she can't get back in'
(CGN session 716)

This example is particularly interesting because the cat in question is referred to by means of masculine pronouns first (by both speakers in the dialogue, including the owner of the cat), before one of the speakers switches to feminine pronouns in the cited utterance. This illustrates that the use of the feminine for female animals is optional, and the masculine may be the more common choice (see also example (8) above).

As regards inanimate entities, only a single instance of a feminine pronoun could be found. The utterance is given in (22).

(22)   dan   kan   ik   je   **dame**   slaan   met   mijn   toren.   maar   ik   kan   je
         then   can   I   your   queen(C)   beat   with   my   castle   but   I   can   you

         ook   dat   we   **ze**   even   wat   beetje   gaan   martelen
         also   that   we   3.F.OBL   PRT   what   little   go   torture

         'Then I can beat your queen with my castle. But then I can... that we'll torture her a bit'
         (CGN session 432)

The referent is the queen in a chess game. In the fictional reality of games, myths or stories, such gender personification is not surprising. The general pattern remains, which is that feminine pronouns are in principle not used in reference to inanimate entities.

5.2.2 Generalization

The above-mentioned uses cover all instances of feminine pronouns in the subcorpus. That is, the body of spontaneous speech investigated in this study only contains a single instance of a feminine pronoun in reference to an inanimate object, which is in all likelihood a case of personification. Thus, there is no evidence of the grammatical, i.e. non-semantic feminine still propagated by the standard dictionaries (e.g. the *Woordenlijst Nederlandse Taal* and the *Van Dale* dictionary). In the spoken standard of nothern Dutch, feminine pronouns are not used for inanimate referents.

Outside the subcorpus defined for this research, another telling example was found. In (23a), the common gender noun *olie* 'oil' is combined with the oblique feminine pronoun *haar*.

(23) a)   deze   **olie**   is   niet   om   in   te   bakken   maar   u   druppelt
            DEM.C   oil(C)   is   not   for   in   to   bake   but   you   sprinkle

            haar   op   uw   salade   uw   pasta   of   waarop   dan   ook.
            3.F.OBL   on   your   salad   your   pasta   or   what_on   then   also

            'this oil is not for frying in, but you sprinkle it over your salads, your pasta or over whatever'
            (CGN session 8032)

The context reveals the source of the feminine: a quote of a written text is blended into the utterance. This is obvious from the polite second-person pronouns *u* (personal pronoun) and *uw* (possessive pronoun), while the speakers address each other informally as *jij* in the rest of the dialogue. What is more, the antecedent

(olijf)olie '(olive)oil' is pronominalized quite differently by the same speakers in (23b) and (c).

	b)	dan	hebben	ze	van	'tzelfde	gebied	ook	de	**olijfolie**.
		then	have	they	from	the_same	region	also	the	olive_oil(C)

		dus	**dat**	komt	uit	'tzelfde	gebied.
		so	DEM.N	comes	from	the_same(N)	region(N)

'then they also have the olive oil from the same region. So that's coming from the same region.'

	c)	't	zit	toch	ook	bij	**olijfolie**	wel	een	beetje	in	hoe	**'t**
		it	sits	AFF	also	with	olive_oil(C)	PRT	a	little	in	how	3.N

		geconserveerd	wordt.	-	ja	'k	weet	ook	niet	precies	hoe
		conserved	becomes	-	yes	I	know	also	not	exactly	how

		ze	**dat**	maken
		they	DEM.N	make

'Because also with olive oil it matters a little how it's preserved. - Yes, I also don't know exactly how they make it.'
(all CGN session 8032)

Apparently, the normal pronoun choice for *olie* 'oil' is the neuter. This usage of the neuter will be discussed in the following section. For now, it is interesting to see that the use of the feminine pronoun for *olie* 'oil' does not extend into the spoken language.

Linking back to the '*haar*-disease', the overuse of the feminine possessive pronoun discussed in section 3.1.6, we see that there is no trace of it in the spoken language sample. This confirms that it is indeed a written-language phenomenon. While a study of Dutch pronoun usage may not neglect it, it will be regarded as only marginally connected to the phenomena discussed in the present study.[2]

5.2.3 Distribution

The corpus data shows that feminine pronouns are distributed semantically rather than syntactically. This generalizations holds for all feminine pronouns in the subcorpus. 40 of these occurred after neuter gender antecedents. As defended in

---

[2] Remember that highly educated speakers sometimes use feminine pronouns when referring to inanimates, especially abstracts, in formal discourse (section 3.1.3). Yet, such usage is not representative for the colloquial, spontaneous speech of average speakers, and it was not encountered in the corpus sample.

section 4.4 above, common gender nouns referring to persons such as *de vrouw - zij* 'the woman - she' were not added to the database due to their lack of variability: speakers invariably choose the feminine pronoun in such cases. The only alternative is the common gender demonstrative with the forms *deze* (proximal) or *die* (distal). The choice between an ordinary personal pronoun and a demonstrative was not part of the present study, having little to do with gender and more with discourse factors such as newness or givenness of information.

(24) Feminine pronouns across semantic classes

Class	Male person/ animal	Female person	Female animal	Animal (general)	Bounded object/ abstract	Spec. mass	Unspec. mass/ abstract
Noun [N]	-	39	-*	-	1	-	-

*but attested elsewhere in the corpus

All 40 instances in the subcorpus refer to persons, with the one exception of *dame* 'queen (in chess)'. Among the pronouns with common gender antecedents, which were not included in the database, inanimate referents were not found either. As for the masculine pronoun, the distribution of the feminine pronoun can be explained in terms of semantics rather than syntax. The semantic patterns here are particularly clear because the feminine is used only in highly restricted environments.

5.3 Neuter

5.3.1 Preliminaries

Neuter pronouns are notoriously difficult to investigate, as they fulfil a variety of functions in the sentence. In the vast majority of cases, neuter pronouns do not have an antecedent or do not refer at all. A search through a random corpus conversation gives the following first ten ocurrences of pronominal *het* (the attributive determiner *het* is disregarded):

(25) maar kijken of **het** ervan komt   'we'll see if that'll happen'
      het wordt een gedoe   'it'll be a hassle'
      het gaat echt makkelijk   'it's really easy'
      hoe ik **het** doe   'how I do it'
      dan vraag ik **het** nog 'ns   'then I'll ask once more'
      dan ga 'k er vanuit dat **het** kan   'then I'll just assume that it's possible'
      het is wel weer aan ja   'they're back together, yes'
                                               (lit.: 'it's on again')

      dan wordt **het** steeds erger   'then it's getting worse and worse'
      anders heeft **het** geen zin   'otherwise it's useless'
      al met al ga ik liever naar wat is   'in all, I rather go to... what is it?'
      het?

None of these examples is suitable for the questions at hand. In most cases, *het* functions as dummy, expletive or temporary subject or object (depending on analysis and terminology). It is often used in existential and predicative constructions, and it participates in a large number of fixed expressions. In none of these circumstances it is referential, unless we count propositions as referents. For the present purposes, only referring instances of *het* with an identifiable antecedent qualify as evidence. This is obvious, as the gender of the antecedent determines whether the pronoun is a match or a switch.

Moreover, neuter pronouns cannot appear after prepositions; instead, the particle *er-* is prefixed to the preposition.

(26) Hij pakte   een **mes**   en ging **ermee** zwaaien
      he grabbed a knife(N) and went with_it wave
      'He grabbed a knife and started waving it about'

In cases like (26), *ermee* replaces *met het 'with it'. Other examples are *ervan* 'of it', *ernaast* 'next to it' *erdoor* 'by/through it', However, *er*-forms are not restricted to neuter gender antecedents. In (26), the common gender noun *hamer* 'hammer' would have been as good an antecedent as the neuter noun *mes* 'knife'. This means that such cases cannot be treated as evidence.

In general, neuter pronouns pose more problems to the analysis than masculine and feminine pronouns, and some grey areas of ambiguity remain.

5.3.2 Use of the neuter pronoun

The neuter pronoun with the full form *het* and the clitic form '*(e)t* appears as a gender switch with a number of common gender nouns. These are - among others - the following: *as* 'ash', *cappucino* 'id.', *grond* 'earth', *huid* 'skin', *kaas* 'cheese', *kip* 'chicken (fowl)', *knoflook* 'garlic', *melk* 'milk', *olijfolie* 'olive oil', *pasta* 'id.', *post* 'mail', *puree* 'id.', *sla* 'salad', *spinazie* 'spinach', *thee* 'tea', *verf* 'paint', *wijn* 'wine', *wol* 'wool' and *zeep* 'soap'. All of these nouns are mass nouns. For a more fine-grained analysis, they were semantically sorted into two classes: specific and unspecific masses. The sorting criteria were outlined in 4.5.4 above. Specific masses are mass nouns that denote specific instantiations, i.e. types, sorts or portions, of substances. Also, mass nouns used with a determiner are put into this class,. By contrast, conceptually unbounded, generic masses and mass nouns without determiner are labelled *unspecific mass*. (27a-b) gives two examples for specific masses, (27b-c) for unspecific masses.

(27) a) een decanteerfles. daar stop je   je **wijn**   in en dan
         a   decanter(C)   there put you your wine(C) in and then

         kan **'t** luchten.
         can 3.N breathe

'A decanter. You put your wine in there and then it can breathe.'
(CGN session 404)

b) die **spinazie** die staat te ontdooien hè. - oh. maar
   DEM.C spinach(C) DEM.C stands to unfreeze hey o but

   dat maakt niks uit als **het** in de koelkast staat
   DEM.N makes nothing out if 3.N in DEF.C fridge(C) stands

   'That spinach is unfreezing, right? - Oh. But it doesn't matter if it's in
   the fridge'
   (CGN session 7811)

c) ik vind **puree** van echte aardappelen altijd lekkerder
   I find puree(C) of real potatoes always tastier

   want **het** is wat steviger
   because 3.N is what firmer

   'I think always puree from real potatoes is better because it is more
   firm'
   (CGN session 683)

d) da's zo handig met **wol** zegt ze dan want je
   that's so handy with wool(C) says 3.F then because you

   kunt **'t** overal tussen stoppen
   can 3.N everywhere between stuff

   'that's so handy about wool, she says then, because you can stuff it
   between everything'
   (CGN session 266)

Looking back at example (23) in section 5.2 above, we see that it reflects the same
pattern. When the speaker switches from reading to spontaneous speech, the
common gender mass noun *(olijf)olie* '(olive) oil' is pronominalized by neuter
pronouns.

Moreover, a neuter personal pronoun appears in three cases with abstract nouns.
These are *taal* 'language (in general)' and *informatie* 'information' (example (28)
and *kunst* 'art'.

(28) m'n oom zou nog **informatie** over reuma doormailen. -
    my uncle would yet information(C) about rheuma through_mail

    en hij had **'t** dus in Word doorgemaild
    and hij had 3.N thus in Word through_mailed

'My uncle was going to send me information about rheuma. - And then he sent it in Word'
(CGN session 400)

There are a few more cases of this sort with neuter gender demonstratives, featuring the nouns *cultuur* 'culture', *muziek* 'music', *kennis* 'knowledge', *spraak* 'speech' and *e-mail* 'e-mail (as a facility)'. These abstracts are different from those listed in 5.1.1, which were *verhaal* 'story', *argument* 'id.', *tentamen* 'exam' and, arguably, *liedje* 'song', *mailtje* '(single) e-mail' and *artikel* 'article'. The neuter abstract nouns pronominalized as masculine are conceptually bounded, they have clear limits and they are countable. By contrast, the common gender abstracts pronominalized as neuter are conceptually unbounded and uncountable. This justifies the choice of sorting the bounded abstracts with the bounded concretes and the unbounded abstracts with the concrete mass nouns rather than assuming a separate class for all abstract nouns.

In a few instances, a neuter gender pronoun appears with a common gender antecedent that refers to an object. Three examples are the following.

(29) a) jullie hebben een nieuwe **auto** hè? - **'t** ziet er netjes uit.
          you have a new car(C) right 3.n looks PRT neat out

       'you've got a new car, haven't you? - it looks neat'
       (CGN session 397)

      b) ''t is mij nog nooit overkomen dat ik een **presentatie**
          it is me yet never happened that I a presentation(C)

          moest inleveren en dat **het** niet af was
          had_to hand_in and that 3.N not finished was

          'It never happened to me that I had to hand in a presentation and that it wasn't finished'
          (CGN session 435)

      c) ze zouden voor allemaal een **propedeuse** invoeren.
          they were_going_to for everybody a bachelor(C) introduce

          het bestond dus nog niet voor Algemene Taalwetenschappen
          3.N existed thus not yet for general linguistics

          they were going to introduce a bachelor degree. That means, it didn't exist yet for General Linguistics'
          (CGN session 336)

Some of these cases (e.g. 29a and b) are probably collocational. In a Google search, we find that neuter gender is dramatically more likely than non-neuter for the different pronouns that can precede the phrase *ziet er mooi uit* 'looks nice':

(30)

Gender	Search string	Number of occurrences
neuter (personal pronoun)	"het ziet er mooi uit"	39.000
neuter (demonstrative)	"dat ziet er mooi uit"	17.100
masculine (personal pronoun)	"hij ziet er mooi uit"	1700
feminine (personal pronoun)	"zij ziet er mooi uit"	11
common (demonstrative)	"die ziet er mooi uit"	187

This may prompt speakers to use the neuter more liberally in this and similar environments, leading to utterances like (29a). The same can also be the reason for the neuter in *het is af* 'it is finished' in (29b).

(29c) is a type of reference that will be discussed in the section on the neuter gender demonstrative (5.5 below). Here, reference is vague and the pronoun is interpreted as 'a thing like that'. This is one of the functions of the neuter in anaphoric reference.

5.3.3 Distribution

Summarizing, in 48 of the 56 common-to-neuter switches, the neuter pronoun refers to a mass entity. The other 8 switches to neuter gender have an object referent. The sample contains no neuter gender pronouns for animate referents (but only common gender antecedents were considered).

(31) Neuter gender pronouns across semantic classes

Class	Male person/ animal	Female person/ animal	Animal (general)	Bounded object/ abstract	Spec. mass	Unspec. mass/ abstract
Noun [C]	-	-	-	8	11	37

Again, the distribution of this pronoun gender lends itself to a description in semantic terms.

5.4 Demonstrative pronouns: common gender

5.4.1 Use of the common gender demonstrative

The common gender demonstrative plays a central role in spoken Dutch. Especially the distal variant *die* is immensely frequent, and it can be used with nearly any noun. As a switch, it occurs in a variety of cases.

In human reference, it can be used both for male and female persons. Again, the neuter noun antecedents show that the pronouns are gender switches.

(32) a) m'n **neefje**   **die**   nam   op
         my   nephew.DIM(N) DEM.C took up
         'my nephew took the phone'
         (CGN session 628)

     b) dat   **meisje**   is dus eenentwintig en **die**   gaat trouwen
         DEM.N girl.DIM(N) is thus twenty-one   and DEM.C goes marry
         'so that girl is twenty-one and she's going to get married'
         (CGN session 631)

     c) 't   **hoofd**   van de productie   als **die**   logistiek onder
         the head(N) of the production if DEM.C logistics onder

         zich   krijgt.
         him/herself gets

         'the head of production, if s/he gets logistics under
         him/her'
         (CGN session 340)

For animals, *die* can be used in all cases; indeed, it can be a handy solution for reference problems when the sex of the animal is not known to the speaker. A typical example is given in (33).

(33) **die**   heeft zoveel   energie dat   **beest**
     DEM.C has   so_much energy DEM.N animal(N)
     'it's got so much energy that animal'
     (CGN session 359)

In the domain of inanimate reference, *die* and *deze* are ubiquitous for objects and, just as we have seen for the masculine in 5.1.2 above, for abstracts of the bounded type. (34) shows two examples each for bounded concrete referents (34a) and abstract referents (34b).

(34) a) **die**   is wel leuk dat   **plaatje**.
         DEM.C is quite nice DEM.N picture.DIM(N)
         'that's a nice one, this picture'
         (CGN session 449)

     b) dit is niet ons **dekbed**. **deze**   is van m'n ouders.
         3.N is not our.N duvet(N) DEM.C is from my parents
         'This isn't our duvet. This one's from my parents.'
         (CGN session 469)

c) een heel sterk **argument**. - **die** telt voor vijf.
a very strong.N argument(N) DEM.C counts for five
'a really strong argument. - That one counts for five.'
(CGN session 9227)

d) je hele **leven die** gaat daar helemaal om draaien
your whole life(N) DEM.C goes there all around turn
'Your whole life that's going to revolve around that'
(CGN session 526)

Moreover, common gender demonstratives occasionally appear with reference to collectives. And example is (35).

(35) maar 't **publiek die** moest betalen.
but DEF.N audience(N) DEM.C must pay
'but the audience, they had to pay'
(CGN session 836)

Collectives are difficult to analyze because they can be conceptualized in different ways, ranging from human agents to quasi-inanimate institutions. The exact intended meaning is difficult to assess, and while pronoun usage is in all likelihood sensitive to such conceptual distinctions, the data does not provide sufficient evidence to base an analysis on. Besides, it should be noted that the preferred pronominalization strategy for collectives is the use of a plural pronoun.

The pattern that generally does not occur is a neuter gender mass noun followed by a common gender demonstrative. The only example is about *stokbrood* 'baguette', which, being food, can be counted as mass, but which is clearly conceptually bounded in its particular shape and also because sorts are being compared.

(36) dat andere **stokbrood** smaakt echt heel sterk naar saté.
DEM.N other baguette(N) tastes really very strongly like satay

is wel lekker hoor maar ik vind **deze** echt veel lekkerder.
is AFF tasty PRT but I find DEM.C really much tastier

'Yes because that other baguette tastes really strongly like satay. It's quite nice, really, but I really think that this one tastes much better.'
(CGN session 468)

Neuter-to-common switches with mass referents are the exception, and there are no cases with true unspecific mass referents.

5.4.2 Distribution

The 145 cases in the database are distributed across the various classes as sketched in (37). Only neuter gender antecedents are considered here.

(37) Common gender demonstratives across semantic classes

Class	Male person	Female person	Human	Animal	Bounded object/ abstract	Spec. mass	Unspec. mass/ abstract
Noun [N]	19	25	1	2	80	2	-

Not in the table are references to events (2 examples), and 9 ambiguous nouns that fit in more than one semantic class. The distribution of the common gender demonstrative combines that of the masculine and that of the feminine pronoun. It has a large range of uses, only excluding unspecific masses and abstracts. Once more, we see that the distribution can be described in semantic rather than syntactic terms.

5.4.3 From demonstrative to personal pronoun

The usage facts show that the common gender demonstrative has all the referring options of masculine and feminine combined. Besides, it functions as a syntactical match to any common gender noun and thus to the majority of nouns in Dutch. Moreover, as noted in 3.1.5 above, speakers employ it as an avoidance strategy when they feel uncertain about the appropriate pronoun gender. All of these factors contribute to the great frequency of this versatile pronoun. In fact, there are reasons to assume that it is losing the discourse functions normally associated with a demonstrative, gradually acquiring the status of an ordinary personal pronoun. A relevant example was mentioned in section 3.1.5 and is repeated here.

(38) De **mummie** zal eerst een **CT-scan** ondergaan voordat **deze**
the mummy(C) will first a CT-scan(C) undergo before DEM.C

tentoongesteld wordt in het Sakkara museum
exhibited becomes in DEF.N Sakkara museum(N)

'The mummy will first undergo a CT-scan before it is exhibited in the
Sakkara museum.'
(daily newspaper *Metro*, 04-05-05)

When two antecedents are available, a personal pronoun is ambiguous as to one or the other. A demonstrative, however, will normally pick out the last of the antecedent candidates.[3] This clashes with the semantics in (38) where the demonstrative must refer to the mummy rather than the scan. An ordinary personal pronoun would not have caused this trouble. While it is still ambiguous and can link to either of the two antecedents, it does not prefer the second over the first, so

---

[3] The resolution preferences of ambiguous demonstrative pronouns can be amplified by contrastive stress. Unfortunately, the example used for illustration comes from a written source, so stress information is not available.

common sense and topic continuity are free to pick out the first as the most likely candidate. However, the personal pronoun poses problems for gender choice, as noted earlier. Officially, *mummie* is feminine (according to the *Groene Boekje*), but this is not part of the competence of the speakers in the north. The spoken language prefers the masculine, but this choice may be felt as too colloquial for a newspaper article. The demonstrative is a safe bet: it sidesteps the decision. If demonstratives are frequently chosen on such grounds, their special antecedent preferences weaken and they develop more resemblance to ordinary personal pronouns.

The development from demonstrative to personal pronoun is in no way unusual. Demonstratives are known to be the ancestors of definite articles as well as of personal pronouns (see Diessel 1999: 8 for references). More specifically, the Scandinavian languages, which in many ways represent a different stage in a similar development as proposed for Dutch, have promoted the former demonstratives *den* [C] and *det* [N] to main anaphoric pronouns (see e.g. Davidson 1990). The original pronouns, e.g. *han* [M] and *hon* [F] in Swedish, are now personal pronouns in the narrow sense: their usage is restricted to human (and occasionally animal) referents. The usage patterns of the common gender demonstrative in Dutch may indicate an early step in the same direction.

5.5 Demonstrative pronouns: neuter gender

5.5.1 Usage of the neuter gender demonstrative

The neuter demonstrative pronouns show a bewildering variety of uses, and careful analysis is needed to separate the different and indeed sometimes contradictory patterns.

In the straightforward instances, neuter gender demonstratives have the same distribution as neuter gender personal pronouns. Thus, they appear as gender switches in combination with common gender mass nouns, either with unspecific (39a) or with specific reference (39b).

(39) a) 't   lijkt   wel   **motorolie**.-   alsof je   **dat**   weleens ophebt
         3.N   seems   PRT   motor_oil(C) as_if you DEM.N ever   on_have
         'That tastes like motor oil. - As if you've ever tasted that!'
         (CGN session 6760)

     c) de   **kerkmuziek**   die   kwam ons keihard   tegemoet
         DEF.C church_music(C) DEM.C   came us   rock_hard   towards

         buiten het   huis.   -   mijn moeder   die   draaide **dat**.
         outside DEF.N house(N)   my   mother(C)   DEM.C   played   DEM.N

         'The church music came blasting towards us outside the house.
         - My mother played that.'
         (CGN session 275)

References to abstract entities are found occasionally, and again, the abstracts are uncountable.

(40) spontane   **spraak**   **dat**   willen   jullie   horen   hè
     spontaneous.C   speech(C)   DEM.N   want   you.PL   hear   right
     'Spontaneous speech, that's what you want to hear, right?'
     (CGN session 609)

Yet, there are a number of cases that do not conform to the usual pattern. They show a *de*-noun referring to a bounded object in combination with a neuter gender demonstrative. In my view, these examples represent a special use of the distal demonstrative *dat*. The issue was already breached above, and we shall briefly discuss it.

5.5.2 Unspecific *dat*

A typical example for this usage of *dat* is given in (41).

(41) **handdoek** hoeft   niet   want   je   zit   in   een   hotel.   als   je
     towel(C)   need   not   because   you   sit   in   a   hotel(N)   if   you

     niet   had   gezeten   had   je   **dat**   wel   moeten   meenemen
     not   had   sat   had   you   DEM.N   AFF   must   take_with

     'a towel isn't necessary because you're in a hotel. If you hadn't been there you would have had to take one with you.'
     (CGN session 439)

The reference here is not to a specific towel, but to a towel in general. There is no meaning change when *dat* is replaced by *zoiets* 'such a thing, a thing like that'.

However, if the antecedent is replaced by a more specific alternative, *dat* is no longer available.

(41)' je   blauwe   handdoek   hoeft   niet   want   je   zit   in   een   hotel.
     your   blue(C)   towel(C)   need   not   because   you   sit   in   a   hotel(N)

     als   je   niet   had   gezeten   had   je   **die/?dat**   wel   moeten   meenemen
     if   you   not   had   sat   had   you   DEM.C/N   AFF   must   take_with

     'Your blue towel isn't necessary because you're in a hotel. If you hadn't been there you would have had to take it with you.'
     (CGN session 439)

A parallel example is (42).

(42) Ik wil   een nieuwe **zomerjas**.   - wat   voor één?
     I   want a   new.C   summer_coat(C)   what for   a

    een lange zwarte? - nee. **dat**   is meer voor de   winter.
    a   long.C black.C   no   DEM.N is more for   DEF.C winter(C)

    'I want a new summer coat. - What sort of coat? A long black one? -
    No. That's more for the winter.'
    (CGN session 469)

The referent is a coat that is unspecified because it has yet to be found and bought. A common or masculine pronoun instead of a neuter would trigger a specific reading not compatible with the utterance semantics. Also, replacement of the pronoun by *zoiets* 'such a thing' is again easily possible. Thus, *dat* is a special anaphor which turns the referent into a representative for a whole class of things.

The same sort of mental move from an individual object to a semantic group can be witnessed in (43). Here *onderbroek* 'underpants' is referred to as *dat*. Later in the utterance, it is joined by two other nouns, *sokken* 'socks' and *t-shirt* (id.), to form the group of items of clothing that should be changed every day. The effect of *dat* is that the underpants are seen as a part of this semantic group rather than as a specific object in the subject's wardrobe.

(43) hoe zit 't dan   met   schone **onderbroek**?   - nou **dat**   trok
     how sits it then with clean.C underpants.SG(C)   now DEM.N pulled

    ie wel iedere dag aan. schone sokken en   schoon t-shirt.
    he AFF every   day on   clean   socks   and clean   t-shirt

    'how about clean underpants? - Well, he did change those every day. Clean
    socks and a clean t-shirt.'
    (CGN session 480)

Similar uses of the neuter occur with plural antecedents and conjuncts. In all of these cases, reference is unspecific and to the pair or group rather than to individual entities.

Note that, as pointed out in Audring (2006a: 100-101), this unspecific *dat* can be used for nouns of any semantic class, even for humans, as long as no specific referent is involved. Compare the constructed examples under (44).

(44) a) Hij heeft een **partner**. **Dat**   had   'ie   eerder niet gehad.
          he   has   a   partner(C) DEM.N had   3.M   earlier not   had
          'He's got a partner. He didn't have one before.'

But:

	b)	Hij	heeft	zijn	**partner**	meegenomen.	***Dat**	heeft	'ie
		he	has	his	partner(C)	with_taken	DEM.N	has	3.M

       daar   ontmoet.
       there   met

       'He's brought his partner. *He's met that there.'

(44a) shows that the function of *dat* is independent of its antecedent noun. Even for common gender nouns with person reference, where a neuter pronoun is normally out of the question (witness 44b), the right (unspecific) context allows the selection of a neuter gender demonstrative. Neuter personal pronouns are not possible in either (44a) or b).

Cornish (1999: 30-31) discusses a similar use of English *that*, which he interprets as a deictic rather than an anaphoric function of the pronoun. His analysis builds on the notion of discourse refocusing: the demonstrative is said to refer not to the discourse referent that is presently in focus, but to bring forward another, previously unfocussed referent. One of the examples given is the following (Cornish 1999: 31).

(45)   Could you send me your **Journal of Semantics** article? We don't have **that** in our library.

Here, the focus moves from *article* to *Journal of Semantics*, and the latter is referred to by the pronoun.

Indeed, it seems that (43) involves refocusing from an individual referent to a group of items, or that in (42), the focus shifts from a single summer coat the speaker wishes to buy to summer coats in general. Yet, it is not clear how this process turns the pronoun from an anaphoric into a deictic device.

Whatever the theoretical decision, such examples should be identified and set aside, as they are more than a mere gender switch. The discourse function of neuter *dat* seems to license patterns that extend the normal semantic range of a neuter pronoun.

5.5.3 Distribution

In general, however, the majority of the examples shows the same semantic pattern as the neuter gender personal pronouns. 43 of the 62 cases refer to mass entities (35 to unspecific, 8 to specific masses). One referent is a place. There are 18 cases with an object or bounded abstract referent. 11 of these are instances of unspecific *dat*, as discussed in section 5.5.2.

(46) Neuter gender demonstratives across semantic classes

Class	Male person/ animal	Female person/ animal	Animal (general)	Bounded object/ abstract	Spec. mass	Unspec. mass/ abstract
Common	-	-	-	18	8	35

The distribution closely resembles that of the neuter personal pronoun. Again, it is clearly semantics-driven.

5.6 Relative pronouns

Relative pronouns are of particular interest for this study because they represent a syntactically different target type, and there are reasons to expect different behaviour for relative pronouns than for personal pronouns. Unfortunately, relative pronouns are quite rare in spontaneous speech: there are only 167 instances in my subcorpus, 22 of which are switches. When relative pronouns are used, they show the same sort of variation as do personal pronouns. Next to the expected matching combinations, common gender relative pronouns can occur with neuter gender antecedents, and vice versa. The latter option - neuter relative pronouns with common gender antecendents - is rare: there are only 4 examples in the database.

Interestingly, the semantic patterns found with switched relative pronouns seem to be the same as for personal pronouns.

Common gender relatives are used for persons, animals and bounded objects. An example for each case is given in (47a-c).

(47) a) een **koorlid** **die** een aanval krijgt
a choir_member(N) REL.C a fit(C) gets
'a choir member that has a fit'
(CGN session 773)

b) een **schaap** **die** mond- en klauwzeer kan krijgen
a sheep(N) REL.C foot- and mouth_disease kan get
'that's a sheep that can get foot and mouth disease'
(CGN session 822)

c) een speciaal **programma** downloaden **die** dat ondersteunt
a special.N program(N) download REL.C that supports
'download a special program that supports this'
(CGN session 400)

The reverse pattern, common gender nouns with neuter gender relatives, is found in reference to mass nouns.

(48) a) **wol** **wat** ze mee moet nemen[4]
wool(C) REL.N she with must take
'wool that she has to take with her'
(CGN session 284)

b) dat er geen **apparatuur** onbeheerd is achtergebleven
that there no equipment(C) unattended is left_behind

dat aan staat
REL.N on stands

'that no equipment is left unattended that's switched on'
(CGN session 252)

Thus, the relatives faithfully match the pattern that we find for the personal pronouns.

There is one case that goes against the general pattern by containing a countable common gender noun and a neuter gender relative pronoun. In all likelihood, the neuter here is triggered by attraction from the intervening neuter gender noun *straatconflict*.

(49) ze hebben ook een uitdrukkelijke **positie** gekozen in
they have also a explicit position(C) chosen in

dat **straatconflict** **wat** niet niet onze positie was
DEM.N street_conflict(N) REL.N not out our position(C) was

'they chose an explicit position in the street conflict that wasn't our position?'
(CGN session 597)

The switched relative pronouns in the subcorpus are distributed among the semantic classes as follows. The table gives all instances, switches from common to neuter gender and from neuter to common.

(50) Relative pronouns across semantic classes

Class	Person	Animal	Bounded object/abstract	Unbounded mass/abstract
Noun [N], pronoun [C]	3	1	14	-
Noun [C], pronoun [N]	-	-	1	3

---

[4] The neuter gender question word *wat* 'what' is frequently used instead of the neuter gender relative pronoun *dat*. This usage is frowned upon in written style.

While the cases are few and far between (see Chapter 8 for frequency data), it is interesting that they do exist, and that their distribution conforms to the patterns observed for personal pronouns. Relative pronouns have particular properties that make their switching behaviour a noteworthy fact. First, it is not immediately clear why they should switch at all. After all, the relative pronouns distinguish two genders - common and neuter - just as the noun and its attributes. If we argue that semantic agreement in pronouns is a consequence of their mismatching paradigms, then relative pronouns do not have a reason to show anything but syntactic agreement. They have the paradigmatic means to agree syntactically with the nouns they relate to, and if we regard syntactic agreement as the norm and semantic agreement as deviant, an explanation is needed for why the deviant option is found in a number of cases. Note that the same reasoning also applies to demonstrative pronouns, which also distinguish common and neuter gender. Second, relative pronouns are typologically less expected than personal pronouns to display semantic agreement. The expectation is confirmed in the corpus data, as will be discussed in Chapter 8. It is also reflected in descriptive grammars, which will usually say that relative pronouns always match the gender of their antecedent noun. These circumstances jusitfy the special attention to the agreement behaviour of relative pronouns.

5.7 Possessive pronouns

There are only very few instances of switched possessive pronouns in the subcorpus. All of them occur after neuter gender nouns with human reference, such as *meisje* 'girl'. For inanimate possessors, possessive pronouns are generally dispreferred in Dutch and are replaced by definite articles. Due to the scarcity of data, this type of pronouns needs to be excluded from the discussion. Yet, more data would have been interesting, as the possessive pronoun in Dutch is exceptional in that it is fully accepted as a semantic rather than a syntactic choice. As already argued in 5.2.1 above, a syntactically agreeing possessive - the neuter *zijn* instead of the feminine '*r* - would be deemed ungrammatical in an utterance such as (51).

(51) **meisje** van   het   PAK   vierde   **'r**   verjaardag.
      girl(N)   from   the   PAK   celebrated   POSS.F   birthday
      'girl from the PAK celebrated her birthday'
      (CGN session 479)

In a typological scale such as the Agreement Hierarchy (Chapter 2), the Dutch possessive pronouns would figure on the right end of the hierarchy, beyond the personal pronouns. It would be interesting to see if this position is valid for other languages.

5.8 Exophoric reference

As discussed in Chapter 1, antecedentless anaphors are a difficult type of evidence, as we have to guess which noun served as tacit antecedent and motivated the choice of one pronoun or another. Yet, if we are only interested in the semantic classes that

are preferred for the referent of a particular noun, exophoric pronouns can be illustrative. In fact, their presumed greater independence from syntactic pressures might allow us to tap into the core semantics of each meaning-based choice. Corbett (1991) attributes the same effect to deictic pronouns: "[t]he semantic content of pronouns is most easily identified when they are used without antecedents, that is, deictically" (244). In Dutch, as in German, the use of deictic pronouns is even more restrictive than that of the anaphoric pronouns, as deictic personal pronouns - excluding demonstratives - are limited to animate referents. An utterance beginning with *hij daar*... 'he there' can only ever be about a person or occasionally an animal. Exophoric pronouns, i.e. anaphors without overt antecedent, are more informative, as they do not have this restriction.

Fortunately, the corpus contains a reasonable number of pronouns without overt antecedent whose referent can be inferred with sufficient certainty. The data shall be reported here in brief.

Almost all of the 80 examples in the database concern references to bounded objects or abstracts, and almost all exophoric pronouns are masculine. This skewing is partly due to the fact that antecedentless neuter pronouns are hard to distinguish from non-referring ones. In particular, reference to taste, smell, look, feel and the like often contain a neuter pronoun, as in the following collocations with the pronoun *het* 'it'.

(52) Het smaakt goed   'It tastes nice'
      Het stinkt   'It stinks'
      Het ziet er goed uit   'It looks good'
      Ik vind het lekker   'I like the taste' (lit. 'I find it tasty')

Since smell, feel, taste and other sensations are often triggered by substances, the neuter pronouns can be instances of exophoric mass reference, but in many cases reference is so vague that it borders on the merely collocational. Therefore, neuter pronouns are difficult to use as evidence for exophoric reference. The only unambiguous case has the implicit antecedent *koffie* 'coffee'.

Antecedentless anaphors can occur when the attention of the speakers is focussed on the same object or circumstance, or when gestures and gaze single out a mutually understood entity that is then available for pronominalization. In the corpus, the referent of exophoric pronouns is often the speech recorder used in the compilation of the corpus. Speakers are meant to ignore the device, which probably prevents them from discussing it openly, but they are often acutely aware of it, especially when unfamiliar with the equipment. The high conceptual salience necessary for pronominal reference is therefore given, and the speakers can reasonably assume that their partner(s) in the dialogue share this state. An example is given in (53).

(53) doet **ie**   't   wel goed? - ja   **hij** doet 't   goed.
      does 3.M 3.N PRT well   yes 3.M does 3.N well

hij neemt gewoon op
3.M takes simply on

'Is it working? - Yes, it's working fine. It's recording normally.'
(CGN session 392)

The referent of the masculine pronouns is a bounded object. This is in line with the observed use of the ordinary masculine anaphor. The other cases confirm this pattern. Example referents are photographs, computers, cameras, books, flowers and telephones. In the example in (54), the referent, the phone, is evoked by the verb *bellen* 'to ring' and is thus eligible for pronoun reference.

(54) ik belde haar eerst maar ja ik denk ik ga **'m** ook niet
      I called her first but yes I think I go 3.M also not

      zesendertig keer laten overgaan
      thirty-six times let go_over

      'I rang her first but well, I think I won't let it ring thirty-six times'
      (CGN session 628)

There are only two instances where an exophoric masculine pronoun refers to a mass noun, in both cases *wijn* 'wine'. The referent is a specific sort of wine that speakers are sampling. Hence, it qualifies for the semantic class of *specific mass*, which has been shown to be eligible for masculine pronominalization. There are no masculine exophors that refer to masses or abstracts.

If we assume that antecedentless anaphors mirror the semantics associated with a particular pronoun, we can conclude that speakers indeed associate masculine gender with bounded objects. Thus, the exophoric data confirms the usage patterns of the anaphoric pronouns. Unfortunately, the difficulties mentioned above do not allow conclusions about the usage of the neuter pronouns.

When the distribution of masculine anaphors and exophors is compared, the question arises why there are no exophoric pronouns for person reference in the data. In my view, this is due to social conventions rather than grammar. The dialogues recorded for the corpus take place indoors, in living room settings where all persons present participate - to varying degrees - in the conversation. In practice, this excludes exophoric reference to persons as it is socioculturally marked to discuss persons within earshot. Exceptions are imaginable in joking, or when the speakers' joint attentions is focussed on a person in a picture or a similar medium. The low likelihood for such a specific scenario makes exophoric pronouns with person reference an improbable phenomenon for a corpus.

5.9 Ellipsis

This chapter concludes with a quick look at another type of data that is interesting for the issues discussed here, though it has nothing to do with pronouns. The relevant construction is an elliptical NP containing a determiner and an adjective, but having its noun elsewhere in the sentence. An example is (55).

(55) 'k wil   wel een **beeld**   maar niet zo'n   **Afrikaanse**
      I   want AFF a   sculpture(N) but   not such_a African.C
      'I do want a sculpture, but not such an African one'

The construction *zo'n Afrikaanse* looks and functions like an attributive to *beeld*. Yet, in normal attributive position it would take the neuter form *zo'n Afrikaans* in agreement with the neuter noun. As attributives agree syntactically in Dutch, it is surprising that ellipses virtually always take the common gender form.

There are reasons to assume that the phenomenon has syntactic rather than semantic causes. When questioned, speakers are uncomfortable with neuter elliptical NPs. They may reason that *een nieuw* is 'more correct' than *een nieuwe* in sentences like (56a), but prefer the latter anyway. The only exception seems to be *een ander* 'another', witness (56b).

(56) a)   Het **glas**   is   kapot? Neem **een**   **?nieuw/nieuwe**
          DEF.N glass(N) is   broken take   a   new.N/ new.C
          'The glass is broken? Take a new one.'

      b)   Het **glas**   is   kapot? Neem **een** **ander/andere**
          DEF.N glass(N) is   broken take   a   other.N/ other.C
          'The glass is broken? Take another.'

In (56b) both variants are accepted (although the latter is considered incorrect in writing). Yet, there are reasons for doubt if *een ander* is actually a neuter gender form. The form also occurs as an independent nominalization, in which case it refers to a person and takes a common gender article, as all other deadjectival nouns with person reference.

(57) de lange   / de blonde   / de nieuwe   / de ander
      DEF.C tall one   DEF.C blonde   DEF.C new one   DEF.C other

Thus, there seems to be a more general syntactic problem with neuter elliptical NPs.

Yet, the issue also has a semantic side. Replacement of a neuter by a common gender form is only possible when the head noun refers to a person, an animal, an object or a bounded abstract entity. When the ellipsis modifies a mass noun, speakers are at a loss.

(58) a) Hebben we nog **water**? Ja   maar geen **?koud/?koude**
          have   we still water(N) yes but   no   cold.N/ cold.C
          'Is there any water left? Yes, but no cold water.'

Neither of the two adjectives feels right. The neuter form is dispreferred in any case, although it might be used in writing, and the common form is unacceptable with mass nouns. At the latter point, common gender ellipses mirror the distributional constraints of common gender pronouns and may be considered gender switches. As to subsequent pronominalization, common gender ellipses are generally followed by common gender pronouns, even if their head noun is neuter. An example is (59):

(59) zo'n   **pak**   zonder stropdas, zo'n   beetje **moderne** en Leonie
      such_a suit(N)   without tie(C)   such_a little modern(C) and Leonie

      zo van ja **deze**   is leuk. en **die**   vond Laurens weer stom
      so of   yes DEM.C is nice. and DEM.C found Laurens again silly

      'a suit without a tie, a bit of a modern thing, and Leonie said, yes, that's a nice one, and Laurens hated it'

This resembles the tendency for pronouns not to reverse switches. Once the agreement has changed from syntactic to semantic, it is unlikely to turn back (see section 8.3 for evidence).

5.10 Conclusion

This concludes the report of the corpus data. We have seen that for each pronoun, a number of semantic classes can be listed that characterize their distribution.

Masculine pronouns are used for

- male humans
- male animals
- non-sex-differentiated animals
- bounded objects and abstracts
- specific masses.

The feminine pronoun occurs with

- female persons
- female animals (occasionally).

Both pronouns occur in these semantic contexts regardless of their antecedents' nominal gender.

The common gender demonstrative is used for all of the above.

Neuter pronouns (personal or demonstrative) are found with

- specific and unspecific mass nouns
- object nouns in unspecific reference (especially when the neuter pronoun is a demonstrative)

This means that pronoun usage can indeed be explained with the help of semantics. The initial expectation proves to be correct. In the following, the usage patterns for each pronoun will be combined into a uniform account that explains the distribution of 'dis-agreeing' pronouns in spoken Dutch. In later chapters, these generalizations will be linked to typological patterns across Germanic and beyond.

Chapter 6

The Semantics
of Dutch Pronominal Gender

In the previous chapter, a number of usage patterns were identified for each pronoun. Most of them confirm earlier observations in the literature, a few are more surprising. Yet others have been observed before, but will be interpreted differently in the following. The aim of this chapter is to compare findings and to provide a unified account in which all the patterns make sense together.

As in the previous chapter, we will temporarily neglect formal motivations for pronoun choice and only look at the patterns of gender switches, i.e. of pronouns that have the 'wrong' gender from the perspective of their antecedent noun. Thus, we will discuss masculine and feminine pronouns for any sort of antecedent, common gender pronouns for neuter antecedents and neuter pronouns for common gender antecedents.

6.1 Animate reference

For the masculine and the feminine pronoun, the patterns are straightforward. There is a correspondence between pronoun gender (masculine, feminine) and natural gender (male, female) which is unanimously supported in the literature. The fact is already mentioned in the grammar by Kruisinga (1924) and by countless other sources. Kruisinga states that "if [neuter nouns] denote persons or animals, [...] the possessive pronouns are always masculine or feminine, and the personal pronouns usually also. The relative is neuter even then." (76). The last point is not entirely confirmed by the new data. There are only 9 instances of relative pronouns with a neuter gender antecedent referring to a person; three of them are switches. (1) is an example.

(1) dat   kleine **meisje die**   'k eigenlijk niet eens ken
     DEM.N little girl(N) REL.C I   actually not even know
     'that little girl that I don't even know, in fact'

Semantic agreement in person reference is acknowledged by the three authoritative sources: the official spelling dictionary (the *Woordenlijst Nederlandse Taal*), the standard reference grammar *Algemene Nederlandse Spraakkunst* (ANS) and the main dictionary, the *Van Dale*. The *Woordenlijst* no longer explicitly motivates its policy on pronoun usage, but older editions state that "words that refer to female persons, even if they belong to the *het*-class, are mostly referred to by *zij*, *ze* and

117

haar" (*Woordenlijst* 1954, cited by Verhoeven 1990: 496, author's translation). The ANS maintains a theoretical split between intratextual and extratextual reference ("binnen- en buitentekstelijke verwijzing", E-ANS § 5·1·2·1), roughly equivalent to the distinction between anaphoric and deictic reference, and states that in extratextual reference, the natural gender of the referent determines pronoun choice. The grammar also acknowledges that neuter nouns can be used with non-neuter (intratextual) anaphors: "reference to a noun such as *(het) meisje* ['the girl'], which always denotes a female person, is [...] almost always with feminine pronouns" (E-ANS § 5·1·2·1b, author's translation). This usage is obligatory for possessive pronouns (E-ANS § 5·1·2·1b). For relative pronouns, it is rarer, but still acknowledged (E-ANS § 5·8·3·2·2b) The same holds mutatis mutandis for male referents. The *Van Dale* dictionary sorts neuter nouns denoting persons into three separate classes: neuter with male reference, neuter with female reference and epicene nouns (which can refer to both men and women). For all of them, non-neuter pronouns are acknowledged. The usage of masculine and feminine pronouns for neuter nouns in person reference is therefore considered as generally known and accepted, albeit discouraged in writing. The corpus data not only confirms the observations, but it suggests that usage of a neuter pronoun in person reference is practically obsolete in spontaneous speech. Even with relative pronouns, the neuter is in danger. As a popular blog puts it (author's emphasis):

(2)

> Dagelijks en overal kom je **het meisje Die** tegen. Je leest en hoort over haar, je verwondert je al niet meer over haar aanwezigheid, je gaat zelf ook over haar spreken. Maar vanmorgen drong de vraag zich ineens op:
> Leeft **het meisje Dat** nog?

> 'Daily and everywhere you meet the girl Who [c]. You read and hear about her, you no longer wonder at her presence, you start talking about her yourself. But this morning the question suddenly arose: Is the girl Who [n] still alive?'
> (http://bieslog.vpro.nl, October 2005)

For other animate referents, the corpus data yields no surprises, either. Animals are generally referred to by masculine or common gender pronouns, with an occasional feminine pronoun for clearly female animals, especially pets. This creates two potentially controversial aspects of pronouns for animals: non-neuter pronouns occur in combination with neuter nouns and masculine pronouns can be used for female animals. The latter has attracted more attention and is discussed quite a bit throughout the literature, among others in Geerts (1966) and (1995a), where it is shown that the phenomenon has been attested at least since the 17th century. It is also mentioned in the ANS, which states that "in the north, [animal names] are generally only treated as feminine when the speaker is aware of the sex of the animal or wants to make it explicit for some reason or other" (E-ANS § 3·3·3·3, author's translation). This means that one can say of a cow that *he* gives good milk, and speakers do not generally see the oddity.

The other aspect, non-neuter pronouns for neuter nouns in reference to animals, has not sparked off much controversy. It is mentioned in several sources, among others in Kruisinga's grammar (1924: 77) which contains the following constructed example:

(3)   Heb je   je   **paard**   verkocht? - Ja, **hij**   werd   me
        have you   your horse(N)   sold   yes 3.M became me

        te   oud. - Hoe oud was **hij/ie** dan?
        too old   - how old was 3.M   then

        'Have you sold your horse. - Yes, he got too old for me. - How old was he then?'

Cross-linguistically speaking, the correlation between the natural gender of humans and animals and the pronominalization of such referents is unsurprising. For Dutch, the influence of natural gender on pronominal reference to persons and animals is obvious and widely acknowledged by linguists and speakers.

6.2 Inanimate reference

There is much less consensus in the field of inanimate reference. Exploration within this area has been slow. Studies have focussed on the contrast between masculine and feminine pronouns, understandably with the question in mind what was happening to the masculine and feminine gender when the determiners were no longer supporting this distinction. This has the drawback that the distribution of the neuter pronouns has not received much attention, or else has been discussed in isolation (e.g. by Romijn 1996). Yet, the use of the masculine and common gender pronouns cannot be understood without a closer look at the neuter.

In the existing literature, as far as inanimate reference is discussed at all, the usual claim is that syntactic agreement prevails. Neuter nouns take neuter pronouns and common gender nouns combine with masculine pronouns. Moreover, some sources, especially reference grammars and dictionaries, predict feminine agreements for certain nouns (see the discussion in 3.1.3). The corpus data shows that neither of the claims is generally correct.

The most eye-catching of the diachronic developments in northern Standard Dutch is the loss of the syntactically motivated feminine pronouns. When southern Dutch speakers use the feminine for nouns such as *muis* 'mouse', they do not intend to convey the meaning that the rodent in question is a female. The ANS gives the following example (author's glossing):

(4)   Als je   die   **muis**   niet kunt vangen, vreet **ze** vannacht
        if   you DEM.C mouse(C) not can   catch   eat   3.F tonight

het	laatste	stukje	op.
DEF.N	last	piece(N)	up

'If you can't catch the mouse it will eat the last piece tonight.'
(ANS § 3·3·3·3)

Moreover, Southern Dutch uses the feminine for inanimate entities. Again, the ANS provides an example:

(5)	Als	de	**tafel**	in	de	weg	staat,	schuif	**ze**	dan	maar	opzij.
	if	DEF.C	table(C)	in	the	way	stands	move	3.F	then	PRT	aside

'If the table is in the way, just move it aside'
(E-ANS § 3·3·3·5)

The spoken data contains no evidence for feminine pronouns in inanimate reference in northern standard Dutch. Even for nouns such as *kunst* 'art', for which both the *Woordenlijst* and the *Van Dale* dictionary give feminine gender, the corpus data yields masculine or neuter pronouns (example (6), the speaker is a teacher in class).

(6)	zij	realiseren	zich	niet	dat	moderne	**kunst**	een	spiegel
	they	realize	themselves	not	that	modern	art(C)	a	mirror

is	van	**z'n**	tijd.
is	of	POSS.M/N	time

'They don't realize that modern art is a mirror of its time'
(CGN session 556)

Also not supported by the corpus data is the observation that feminine pronouns occasionally occur with mass nouns. This usage is reported in Kruisinga (1924: 76), in Van Haeringen (1936: 21 and 1954: 3) and again more recently in Maljaars (1979: 107) and De Vries (2001: 101), who sees it as a relict from an earlier stage in the development of the language. Maljaars maintains that the feminine has assumed a connotation not in terms of biological female sex but "an emotional value [...] of the collective, abstract, non-concrete, uncountable, vague" (122, author's translation). He concludes that nowadays the feminine has an "abstracting" rather than a "feminizing" function (122). While such a view has occasionally been voiced for the feminine (nominal) gender in Germanic (e.g. in Vogel 2000 for Modern German), the spontaneous spoken data of modern Dutch provides no reason to assume a link between abstractness and feminine gender for this language, as all feminine pronouns refer to female persons or animals.

Moreover, the extensive use of the feminine possessive *haar* for collectives and inanimate entities (see 3.1.6), much discussed among the critical public, is not found in the spoken data.

The fact that pronominalization of formerly feminine nouns has been taken over by other pronouns, notably the masculine, has led many researchers to the conclusion that all non-neuter gender nouns that do not explicitly denote female persons or animals are in fact masculine.

As a statement about nouns, this claim leads to terminological confusion since the proposed masculine gender is not visible on the nouns themselves nor on the attributive elements. As a consequence, one occasionally reads statements such as the following: "Nouns of common attributive gender are feminine when they denote female persons; sometimes also when they denote female animals, but only when the sex is insisted on. All other nouns of common gender are masculine." (Kruisinga 1924: 75). Clearly, a noun cannot be common and masculine or feminine at the same time.

More descriptive and theoretically less controversial is the same generalization in a different guise: formulated in terms of pronouns distribution rather than as a claim about the gender of the noun. An example is Geerts (1995a) who discusses *de*-words (i.e. common gender nouns) and maintains that *hij* and *hem* can refer to all nouns except those referring to female persons (46). The same idea is voiced in Kruisinga (1924), Dekeyser (1980), Hoppenbrouwers (1983), Geeraerts (1992), Geerts (1966, 1995c) and Van der Sijs (2004). In addition, the standard reference grammar embraces this view: "With respect to *de*-words that do not denote persons or animals the following holds. In the spoken language of the north, these words are generally treated as masculine." (E-ANS § 3·3·3·4, author's translation).

The idea seems to be that there is a syntactic mapping between common nominal gender and masculine pronominal gender. This means that common gender nouns with female referents take feminine pronouns by a semantic rule, all other common gender nouns take masculine pronouns by a syntactic rule, and neuter nouns take neuter pronouns, again by a syntactic rule. In this scenario, Dutch has two grammatical genders: masculine and neuter, and one semantic gender: the feminine (although many studies would count the masculine pronoun in person reference as an instance of another semantic gender).

The corpus data reported in the present study suggests a different scenario. On the empirical side, the data shows a class of common gender nouns for which the masculine pronoun is **not** used. These are the unbounded/generic masses and abstracts. Masculine pronouns for generic mass referents are dispreferred so strongly that they are not even used in writing. A sentence such as (7) is judged unacceptable by native speakers, and it has no parallels in the corpus data (the possessive is fine, as it is syncretic with the neuter form).

(7) **Groente**   moet je   niet te   lang koken omdat   ***hij** anders
vegetables.C must you not too long cook   because 3.M otherwise

zijn   voedingsstoffen verliest
POSS.M/N nutrients   loses

'You mustn't cook vegetables too long or they will lose their nutrients'

De Vries (2001: 100) and Romijn (1996: 38) give other examples for *de*-words that resist masculine pronominalization, these are *suiker* 'sugar', *rijst* 'rice', *sneeuw* 'snow', *mosterd* 'mustard' and *wijn* 'wine', all used as mass nouns. The unacceptability of such *de-hij* combinations contradicts accounts that propose a syntactic rule behind the pattern. In Chapter 5 above, it has been argued that the motivations are indeed semantic rather than syntactic. This evidence testifies against the assumption that Dutch still possesses a masculine grammatical gender, i.e. a group of nouns not defined by any semantic commonality but only by the fact that they are pronominalized with a masculine form. The corpus data shows that such a group, when it exists, fails to include all *de*-nouns (even subtracting nouns referring to female persons is not sufficient to define such a group). This does not support an analysis in terms of syntactic mapping.

One might resort to an analysis where masculine agreement is triggered automatically for *de*-words, and that in cases like (7) semantics merely overrides this choice. Yet, the masculine pronoun is also actively used beyond the domain of *de*-nouns: it is regularly found with neuter antecedents.

This usage and its semantic background features occasionally in the literature. An influential account is Van Haeringen (1936), who observes and recommends in a report to the Dutch Ministry of Education that any object ("voorwerp", thing) may be pronominalized by the masculine pronoun (1936: 20). In another book, devoted entirely to the question of pronoun usage in spoken northern Dutch, the same author explicitly says that this use of the masculine extends to **neuter** object nouns and is, in fact, "rather frequent" (Van Haeringen 1954: 15-16). Yet, this usage of the masculine in combination with inanimate neuter gender antecedents is not generally known, or is regarded as substandard. Considering its remarkable frequency in the corpus data (see Chapter 8), it has so far not received the attention it deserves.

The corpus facts indicate that the masculine is a 'semantic gender' in much the same way as the feminine, the only difference being the wider extension of the class, from masculine persons across animals to inanimate objects.

Meanwhile, the developments have not bypassed the neuter gender. While masculine pronouns are invading the domain of the neuter, neuter pronouns in turn appear as anaphors to common gender nouns. The neuter pronoun is the preferred pronoun for mass noun antecedents. This usage is also acknowledged by Van Haeringen (1954: 16) and Maljaars (1979: 14, 105). There is also a more recent

study focusing entirely on the neuter pronoun *het* and containing interesting data (Romijn 1996). The author reports that neuter pronouns can occur with non-neuter antecedents when these refer to substances or 'heterogeneous collections' (an example is *rommel* 'junk'), but also to abstracts such as *haat* 'hate', *pijn* 'pain', *kou* 'cold' or *waarheid* 'truth'. All cited impressions are corroborated by the present corpus study, with the proviso that abstracts are only pronominalized with a neuter pronoun when they are conceptually unbounded or uncountable. Countable abstracts such as *naam* 'name', *bekeuring* 'parking ticket, fine', *advertentie* 'advertisement' or *dag* 'day' take masculine or common gender pronouns.[1] Generally, the corpus data confirms that all gender switching is motivated by semantic reasons.

So far, the discussion has only touched on individual usage patterns. The next step is to systematize and generalize across the observed tendencies.

6.3 A unified analysis

From the usage patterns discussed above, we can distil two parameters that influence pronoun choice. The first parameter is **natural gender** with the values [male] and [female]. As in many gender systems, the masculine and feminine pronouns are used for male respectively female persons and occasionally animals. The second parameter, which governs the usage of pronouns in inanimate reference, appears under different names in the literature. The best terms are **boundedness** or **countability**, and the values are [bounded]/[unbounded] or [count]/[mass]. The split is that between "discrete entities with a well-defined shape and precise limits" on the one hand and "homogenous undifferentiated stuff without any certain shape or precise limits" on the other (Koptjevskaja-Tamm 2000: 1067). In Dutch, pronominalization seems to be sensitive to this parameter. Speakers prefer the masculine or common gender pronoun for bounded, countable entities and the neuter pronoun for unbounded, mass referents.

There are only three accounts that acknowledge boundedness or countability as the main influence on anaphor choice in reference to inanimates. These are Fletcher (1987), Verhoeven (1990) and De Vries (2001).

Verhoeven (1990) explains pronoun usage in spoken Dutch with the help of the following five parameters:

[± unique]
[± bounded]
[± human]
[± feminine]
[± collective]

---

[1] Countable abstract nouns such as *dag* 'day' or *bezoek* 'visit' are occasionally pronominalized by a neuter pronoun when they are construed as events: the standard subject pronoun for *duren* 'take time' is the neuter *het* (*het duurde lang* 'it took a long time').

Excluding proper names ([+unique]) and collectives ([+collective]), which were not considered in the present study, these parameters match those that emerged from the corpus data. Similar observations are made in De Vries (2001), a refreshingly courageous critique of the artificiality of written language norms. A weakness of this account is that it invokes the factor countability only for abstract nouns, failing to show that the masculine for concrete object nouns and the neuter for mass nouns are also two sides of the countability coin. The best account, though little known and in fact unacknowledged by the two others, is a short publication by Fletcher (1987). This study takes the generalization a step further by proposing a factor that is superordinate to all parameters: [count] and [mass], [general] and [specific], [animate] and [inanimate] and [human]/[nonhuman]. This factor he calls "relative degree of salience as an individual" (Fletcher 1987: 62).

Taking up from here, the present study proposes that the semantics of Dutch pronoun gender can be expressed on a conceptual hierarchy which may be called ***Individuation Hierarchy***. This is a variant of the Animacy Hierarchy, a scale commonly ascribed to Silverstein (1976) and widely applied in typological research. Some prominent variants are Givon's Topicality Hierarchy (1979), Dixon's 'potentiality of agency' scale (1979), Sasse's Continuum of Individuality (1993: 659) and Lehmann's Empathy Hierarchy (Lehmann 1988, cf. Kuno 1977).

The Individuation Hierarchy used here has persons as the highest class, separated into male and female (contrary to some versions of the Animacy Hierarchy, local person and third person pronouns are not distinguished, nor are pronouns from nouns. Proper names are not considered, and "kin" is not regarded as a separate class). On the right side, where it usually ends with the class "inanimate", the scale is subdivided into objects, specific masses and unspecific masses. This continuation is entirely in the spirit of the original scale, as will be argued below. Last, but not least, abstracts are not considered a separate class but divided into bounded and unbounded and sorted with their concrete counterparts, as defended in section 4.5.4 above.

The resulting hierarchy is given in (8), with an example noun (phrase) for each class.

(8) Individuation Hierarchy

male human / female human	>	animal	>	bounded object/abstract	>	specific mass	>	unspecific mass/abstract
father sister		*sheep*		*book/name*		*this tea*		*sand/growth*

The difference between specific and unspecific masses cannot be expressed in simple nouns, as most mass nouns can assume a count reading in the relevant

context. Such contexts can be sorting, portioning, comparison (presupposing sorting or portioning), pluralization and others. In (8) above, the unspecific masses/abstracts are illustrated by singularia tantum, while the specific mass is represented by a definite noun phrase.

The conceptual property that unites the elements on the hierarchy is not that of animacy, as animacy is only relevant to the left side of the scale. A closer fit is achieved by taking *individuation* as the basic property that holds the scale together. As the terminological field is wide and disorderly, this decision needs some motivation.

With Fraurud (1996), individuation is understood as a property assigned to entities according to the human world view. It rests on "[...] an anthropocentric cognitive ontology, which is structured around ourselves and our fellow human beings, and where everything else is described from the point of view of human beings" (67). Thus, it is a property derived from actual physical characteristics of a referent - for those referents that are physical objects - and its perception by and in relation to the speaker. In other words, individuation is a property derived from both empiry and empathy, where higher empathy correlates with higher individuation.

From this it follows that referents are most highly individuated when they are adult persons, and that individuation decreases with greater conceptual distance to this referent point. Roughly, humans are more highly individuated than animals, animals more than inanimate objects and those more than substances or uncountable abstracts.[2] This is also the idea behind the Animacy Hierarchy. The present understanding of individuation also captures more subtle semantic differences within and between the groups, smoothing out the gradations of the scale. For human referents, adults are more highly individuated than children. This difference reflects those properties that in human conceptualization distinguish humans and animals: rationality and sentience. Animals are less individuated than humans, with the so-called higher animals at the more individuated end and the 'lower' animals at the less individuated end. Next follow inanimate entities with distinct shapes, typically concrete, middle-sized objects whose boundaries are salient in their conceptualization.[3] Again, among the objects, we may single out a group of referents

---

[2] As Bechert (1982: 23) puts it, "[w]ithin the animacy hierarchy *male persons* are superior to non-male persons, *adult persons* are superior to non-adult persons, *persons* are superior to non-persons, e.g. to animals, *animate beings* are superior to inanimate things, *inanimate things* may be differentiated into solid countable objects and the rest; further divisions are possible."

[3] The condition that boundaries must be conceptually salient is inspired by Jackendoff (1991) who notes that the boundedness of objects and events depends on construal. Unbounded entities and states mass nouns and ongoing events are construed with their boundaries "not in view or out of concern" (1991: 18). This view is similar to Allan's (1980) account of count-mass construal. Jackendoff's approach caters for a linguistic and an empirical fact. The (trivial) empirical fact is

that is higher in individuation than the rest. This is the semantic class of machines and intelligent devices, such as cars or computers, artefacts that can perform tasks partly or wholly on their own. The high individuation of such entities is intuitively expressed in their ability to bear proper names (and in the quasi-social relationships that people may entertain with them). They are often metaphorically regarded as animates and they lend themselves to personification.

At the lowest end of the Individuation Hierarchy are those entities that are conceptualized with loose or no boundaries. These are the typical 'mass' entities such as liquids and other substances. Their lack of conceptual boundaries coincides with non-countability.

As mentioned earlier, the present use of the concept individuation involves the split of abstract referents into bounded and unbounded. Other approaches regard concretes as inherently more individuated than abstracts (e.g. Hopper and Thompson 1980: 253, Timberlake 1975), to the extreme that it is claimed that "[a]bstract nouns refer to concepts which inherently cannot be individuated" (Timberlake 1975: 124). Yet, there are intuitive differences in individuation between abstract referents, roughly parallel to the differences in the countability of abstract nouns. A word is a conceptually bounded, countable and thus more individuated entity than, say, friendliness, which is unbounded and uncountable. Thus, we follow Langacker (1991: 63 ff) and assume that boundedness is not limited to the physical domain. As Koptjevskaja-Tamm puts it: "Bounding can apply to various domains - a cow, as a physical entity, has spatial limits, a beep is bound both in time and pitch and a chapter is bound within a written work" (2000: 1068).

Last, but not least, individuation as understood here can capture other semantic differences such as those between definite and indefinite and those between generic and specific reference. These differences can apply on all levels of the hierarchy. A person or object that is identifiable among other persons or objects denoted by the same noun is more highly individuated than a person or object whose exact identity is vague. The difference between generic and specific is best seen in the context of mass reference. The purest use of a mass noun is in generic reference, and generic mass nouns are the least individuated of all. Thus, there is a difference in individuation between the two uses of *beef* in (9).

(9) a) The beef is too spicy.
     b) I don't eat beef, I'm a vegetarian.

---

that virtually all entities are bounded, so that unboundedness is in most cases a product of the speaker's mind, a consequence of disregarded boundaries. The linguistic fact is that the referents of many nouns can be construed as either count or mass entities respectively with or without reference to their boundaries.

Dutch pronominalization is sensitive to such distinctions. If the patterns of pronoun usage are aligned to the hierarchy, a clear picture results that provides a uniform key to the data. Schematically, Dutch pronoun usage is organized as follows.

(10) Dutch pronominal gender and the Individuation Hierarchy

Semantic class	female human / male human	>	animal	>	bounded object/ abstract	>	specific mass	>	unspecific mass/ abstract
Personal pronoun	feminine								
			masculine				neuter		
	masculine								
Dem. pronoun	common								
			common				neuter		
	common								

The schema shows that alignment to the hierarchy combines all of the usage facts into a logical pattern. Each pronoun has its own domain on the hierarchy. The feminine aligns with the property [female human] (with a small extension to the right for female animals), the masculine with anything from male humans to specific masses, while the neuter is found at the right end of the scale, for specific and unspecific masses. Common gender demonstratives combine the domains of masculine and feminine personal pronouns, and neuter gender demonstratives line up with neuter personal pronouns. In this way, the hierarchy serves as a ***semantic map*** (Anderson 1982, Bybee 1985: 195 f, Croft et al. 1987, Haspelmath 1997) that accommodates all of the seemingly disconnected usage facts.

This analysis of the situation has three other major benefits. It

- accommodates the variation
- is plausible in a context of language development
- makes sense typologically.

We will briefly discuss the three points in turn.

The most remarkable characteristic of Dutch gender agreement is its amount of variation. Take a noun such as *vis* [C] 'fish'. Syntactically, it takes common gender agreement. Its semantics, though, can trigger masculine, feminine and neuter agreements. This clashes sharply with the standard expectation that lexical gender is fixed and agreements are consistent. Yet, all three scenarios can be accommodated within the hierarchy. As an animal, fish triggers masculine agreements. For an ichthyologist or a pet owner, a fish may be a female individual they wish to address with a feminine pronoun. Else, fish can be downgraded to a substance, a type of

food or an ingredient of a meal. This yields neuter agreements. The main point is that the hierarchy is not so much about noun semantics, but about reference in context. Depending on construal, the referent of a noun phrase can appear higher or lower on the scale.[4]

The second major advantage of the proposed account is that it is easy to imagine how such a semantic system could arise and take this particular shape. Cognitive hierarchies of the mentioned type are thought to be universal. Where a syntactic rule ceases to regulate a particular grammatical process, basic conceptual distinctions such as rationality, animacy or countability can step in to the fill the gap. When this happens, we expect a process of polarization which after some time results in a division of labour between morphemes, words or constructions (in our case, pronouns). Positive motivations for using a particular pronoun are reinforced by reasons for avoiding another. A major factor is ambiguity. Consider (11) below (from an e-mail conversation):

(11) ik dacht   dat jouw cadeaubon   naar mijn emailadres
      I   thought that your gift_coupon(C) to   my   e-mail_address(N)

      verzonden was, maar **dat**   klopte   niet
      sent   was   but   DEM.N   was_right   not

      'I thought your gift voucher was sent to my e-mail address, but that wasn't right'

In this example, the pronoun *dat* 'that' can refer to the proposition 'your gift voucher was sent to my e-mail address' or else to the entity 'e-mail address'. In both cases, a neuter pronoun is appropriate, as *emailadres* is a neuter noun and discourse deixis is realized by means of a neuter pronoun (often, though not exclusively, a demonstrative). The availability of an alternative is an attractive option. An e-mail address is an abstract entity, but countable, and thus eligible for a masculine or common gender pronoun. A speaker wishing to make sure that reference is to this entity rather than to the whole proposition can opt for (11)' below (although then ambiguity may arise with the other noun in the sentence).

(11)' ik dacht   dat jouw cadeaubon   naar mijn emailadres
       I   thought that your gift_coupon(C) to   my   e-mail_address(N)

       verzonden was, maar **die**   klopte   niet
       sent   was   but   DEM.C   was_right   not

       'I thought your gift voucher was sent to my e-mail address, but that wasn't right'

---

[4] The term "construal" is borrowed from Langacker (2008: 55 ff). Here, it is used for alternate ways of conceiving and portraying a referent.

If such choices pile up, speakers may become hesitant about using a neuter pronoun for object reference. The neuter then becomes more strongly associated with propositions and masses. Thus, ambiguity avoidance can consolidate the functional divide between the pronouns. However, in a language with only two nominal genders, chances are high that two competing antecedents have the same gender. In this light, pronoun resolution cannot be expected to rely too heavily on feature matches.

The division of labour between the genders leads to the third point in favour of the analysis proposed in this section. This is its typological plausibility. The alignment of the genders to the hierarchy produces a picture in which each pronoun is associated with a single domain on the scale. In other words, the semantic classes for which the same pronoun can be used are adjacent to each other. This is a typologically expected situation which has been described by Haspelmath (1997: 62) and formulated under the name *Semantic Map Connectivity Hypothesis* in Croft (2001, 2004). Siemund (2008) describes the alternative: "We would expect no variety of English to use, say, masculine pronouns for male humans and liquids while using neuter pronouns for animals and countables" (Siemund 2008: 4). The 'disagreeing' pronouns in Dutch comply with the expected patterns.

Note that there are two usage patterns that contradict the Connectivity Hypothesis. The first is the feminine gender for mass referents, mentioned by Kruisinga (1924: 76), Van Haeringen (1936: 21 and 1954: 3), Maljaars (1979: 107). According to De Vries (2001: 101) and the modern corpus data, this usage has disappeared from the system. We have no proof that its limited life-span is related to its involving a disconnected reference domain, but such a hypothesis is not implausible. Second, the use of the feminine possessive *haar* for object nouns and collectives does not fit the proposed pattern. Yet, we argued that it is a written-language phenomenon driven largely by system-external factors such as hypercorrection and stylistic wishfulness.

The issue of connectivity will be of interest in Chapter 10 where pronominal gender languages are discussed.

In stressing the typological expectedness of the patterns of pronoun usage in Dutch, we have sidestepped a point in which the Dutch pronominal gender system does offer reasons for surprise. This is the fact that it employs countability as a main parameter.

6.4 Countability and Gender

Boundedness or countability as a conceptual property is regularly reflected in the grammars of the world's languages, although it usually manifests itself in the context of number (see e.g. Corbett 2000: 96 ff for an overview of the relevant literature). There are clear parallels between mass nouns and plurals, both in terms of semantics and in terms of syntactic distribution (e.g. Allan 1980, Chierchia 1998). However, the link between countability and gender is rarely made. Exceptions can mainly be

found in the Indo-Europeanist research tradition, which has frequently pointed out correlations between mass nouns and neuter gender. For Proto-Indoeuropean, some scholars claim that: "[o]n the lowest end of the animacy hierarchy, we consistently find neuter nouns: these include all of the nouns for masses and fluids" (Matasović 2004: 134).

Countability is not listed among the criteria on which semantic gender systems are based (de la Grasserie 1898 and Corbett 1991: 30-32). Corbett's typology does not mention a gender system in which countability is the defining criterion of a gender, much less a system resting entirely on a count-mass distinction. In fact, such a system would be most problematic, as countability is a construal-dependent property, which means that most nouns would constantly oscillate between two genders. Yet, consistent count-mass splits in gender systems have been observed. Siemund (2008) discusses varieties of English where pronoun usage is sensitive to countability. These are dialects in the southwest of England, the varieties Newfoundland English and Tasmanian Vernacular English (Pawley 2002, 2004), and there are traces of it even in informal spoken American English (see also Wagner 2003). Beyond English, Fernández-Ordoñez (to appear) reports relevant patterns for varieties of Danish, and there is interesting evidence from dialects of Central Italy (Haase 2000) and Spain (Lüdtke 2001, Viejo 2001, Fernández-Ordoñez 2006-2007). Outside Europe, influences of countability have been found in the Bantu languages (Denny and Creider 1976), the Yeniseian language Ket (Werner 1997) and the two Nakh-Daghestanian languages Lak and Archi (Corbett 1991 and sources there). Also, there are some interesting patterns of gender marking connected to individuality/collectivity in the Gulf language Tunica (Haas 1940) and in Arabic (Holes 1994, Hämeen-Anttila 2000).

Especially interesting among the cases discussed in Siemund (2008) are the Wendland dialect of Lower German, described by Rohdenburg (2004a,b), and West Jutish, a variety of Danish described in Ringgaard (1973), Wahrig-Burfeind (1989), (Gachelin 1991) and Allan et al. (2000).

In the dialect of Wendland, belonging to the Lower German language area between Hamburg and Hannover, there is some interesting evidence for count-mass sensitive gender agremeent. It operates not only on the pronominal level, but also on the definite article. Rohdenburg (2004a: 347) gives examples in which a noun has masculine gender when it denotes an object (12a), while the same noun as the head of a compound is neuter when referring to a substance (12b).[5]

---

[5] Rohdenburg (2004a,b) calls the non-neuter gender 'masculine'. However, the definite article *de* that characterizes this gender is also used for nouns that are feminine in Standard German. As there are no other attributive differences between masculine and feminine (Rohdenburg 2004b: 98), it is better glossed as *common gender*.

(12) Wendland dialect

    a)   As   ick denn andol köm un Oma   **de**   **Worst**
          when I   then down came and grandma DEF.C sausage

          heenholle
          offer

          'When I came downstairs and offered grandma the sausage'

    b)   Hol doch mol ju   Mudder, se   schall mol   proben, ow
          get prt   once your mother she should once try   whether

          an **dat**   **Mettworst** noch Peper an   mütt
          with DEF.N mettwurst   yet   pepper with must

          'Go and get your mother, she should try if the mettwurst (type of sausage) needs more pepper.'

A similarly fascinating pattern can be seen in West Jutish, a variety of Danish spoken on Denmark's mainland. This language distinguishes between count and mass referents and marks the split on pronouns and articles (the pronoun have two additional forms for masculine and feminine, Wahrig-Burfeind 1989: 283). Count nouns take common gender agreement (*den æg* 'the.C egg(C)', *den træ* 'the.C tree(C)'), while mass nouns are neuter across the board (*det sne* 'the.N snow(N)', *det regn* 'the.N rain(N)') (Allan et al. 2000: 20). It is claimed that the agreements change depending on context and construal (examples after Ringgaard 1973: 31, author's glossing). Compare (13a-b) and (14a-b).

(13) a)   Æ   **egetræ den**   er stor
          DEF oak tree DEM.C is big
          'The oak tree is tall'

      b)   **Egetræ det**   er bedst til   møbler
          oak tree DEM.N is best   for furniture
          'Oak wood is best for furniture'

(14) a)   **Den**   **fisk** a fanget i   søndags
          DEM.C fish I caught on Sunday
          'I caught that fish on Sunday'

      b)   al **det**   **fisk** a ku   spis
          all DEM.N fish I can eat
          'all the fish I can eat'

The correspondence between countability and gender agreement is so strong that Wahrig-Burfeind (1989: 256) speaks of a number rather than a gender system.

Support for this analysis comes from the neuter pronoun which is increasingly used for plural referents. Such a system is exceptional and begs closer investigation.

The West Jutish genders have much in common with the Dutch system, not only semantically, but also in their developmental pathway. According to the sources, West Jutish has lost gender agreement in the attributive domain (Bechert 1982, Braunmüller 2000; Ringgaard 1973, Wahrig-Burfeind 1989). It was then reintroduced though the determiners, which function as anaphors and as attributive demonstratives. These forms seem to have brought the count-mass pattern from the pronominal into the attributive domain.

For Dutch, there are good reasons to assume that the count-mass split in the genders has also originated in the personal pronouns. Crucially, attributive gender is not sensitive to countability in any obvious or systematic way. Thus, the semantic gender system sketched in this chapter can be seen as a case of ***resemanticization*** (term from Wurzel 1986) of the pronominal genders.

6.5 Resemanticization

The semantic system that we see in the distribution of the pronouns is a relatively recent phenomenon. Combinations of neuter nouns like *boek* 'book' and non-neuter pronouns such as *die* 'that' are more common for young speakers than for their parents (see Chapter 8). Importantly, the patterns of pronominalization have no counterpart within the noun phrase. Exceptions are a handful of double gender nouns such as *de diamant* [C] 'the diamond (stone)' vs. *het diamant* [N] 'the (substance) diamant' or *de haar* [C] 'the (single) hair' vs. *het haar* [N] 'the (mass of) hair'. Here, the determiner mirrors a count/mass split. Yet, in other pairs only one variant is specialized for count or mass reference. Thus, *de steen* [C] 'the stone' refers to a piece of stone as well as to stone-as-material, while *het steen* [N] 'the stone' is necessarily the material. Conversely, *de doek* [C] is a piece of cloth and thus countable, while *het doek* [N] translates as the mass noun 'fabric' but also as the count noun '(movie) screen, canvas, curtain'. A similar case is *kurk* 'cork'. In the meaning of 'bottle seal', *cork* has common gender, while it can have both genders in the mass reading. A large number of mass nouns can be *de* or *het* without difference in meaning (E-ANS 3·3·2·4·i·2). Aside from such doublets, there are plenty of nouns for which the correlation does not hold. Common gender mass nouns are, for example, *honing* 'honey', *boter* 'butter', *koffie* 'coffee', *thee* 'tea', *melk* 'milk', *suiker* 'sugar', *peper* 'pepper', *wol* 'wool', *olie* 'oil', *was* 'wax', *benzine* 'petrol', *kleding* 'clothing', *gel* 'id.', *make-up* 'id.', *shampoo* 'id.'. This group also contains relatively recent loanwords (the last three examples), where we would expect to see the principle at work, if it were part of the assignment system. Finally, some suffixes used to derive (mostly) uncountable abstracts are associated with common gender. Examples are *-age* (as in *spionage* [C] 'espionage'), *-heid* (as in *blijheid* [C] 'gladness'), *-te* (as in *hitte* [C] 'heat'). Thus, we do not have good evidence that there is a gender assignment principle that attributes common gender to count nouns and

neuter gender to mass nouns. This suggests that the semantic gender system is an innovation from pronominalization.

Following Wurzel (1986), the term ***resemanticization*** is proposed for this new functionality of the personal pronouns. While in the past the feminine gender of *tafel* 'table' and the masculine gender of *stoel* 'chair' carried no semantic associations, not even in pronominalization, feminine pronouns now signal (female) personhood and masculine pronouns indicate that the referent is male or countable. This semanticity is clearest in cases where the syntactic gender is felt to be inappropriate.

(15) a) **Vis**   bevat   niet alleen gezonde vetten, **het/?hij/?zij/?die** is
           fish(C) contains not   only   healthy   fat   3.N/M/F/C   is

           ook een leverancier van vitamine D.
           also a   deliverer   of   vitamin D

           'Fish doesn't only contain healthy fat, it's also a source of vitamin D'
           (magazine *Margriet* 11/2008)

      b) 'k wou   m'n **broertje**   bellen of **ie/?het** ook zin   had
           I   wanted my   brother.DIM(N) call   if   3.M/N   also desire had
           'I wanted to call my brother to see if he was in the mood'
           (CGN session 411)

Examples such as these prove the loss of semantic innocence of pronominal gender agreement. At the same time, they show that the Individuation Hierarchy makes the correct predictions.

Within the context of the Dutch gender system, resemanticization can be seen as a reaction to the eroded gender distinctions in the attributive domain. When speakers lost the knowledge about the former masculine and feminine gender, they also lost the original system that governed pronoun usage. This opened the gates to redistribution of the pronominal genders. If the new system of pronoun usage is indeed a reaction to the historical problem, then it constitutes an interesting case of self-regulation within the language system. The semantic rules are typologically logical and natural, and they provide a solid basis for pronoun usage in a mismatched gender system.

6.6 Conclusion

In this chapter, the observations from the corpus study are compared to earlier findings in linguistic literature. It is shown that two usage patterns are of particular interest. First, there is the use of the masculine and common gender pronouns for objects and bounded abstracts, even when the antecedent noun has neuter gender. Second, mass nouns of common gender are often pronominalized with a neuter pronoun. Both tendencies suggest that Dutch pronoun gender is sensitive to boundedness, countability or individuation of the referent.

The usage patterns are projected onto a typological scale called the Individuation Hierarchy. For each pronoun, a semantic domain is identified that characterizes its function. Interestingly, each domain represents one connected area on the scale, suggesting a division of labour among the pronominal genders. This state of affairs is typologically expected and predicted by the Semantic Map Connectivity Hypothesis.

The analysis in terms of a semantic hierarchy is covering all the relevant facts in a unified analysis. The quirky Dutch use of the 'wrong' pronouns is thus shown to be systematic, and is linked to typologically familiar behaviour. At the same time, the account gives theoretical room to the observed variation.

The last point, variation, is the topic of the following two chapters. While Chapters 5 and 6 have focussed on switched pronouns only, Chapters 7 and 8 expand the focus towards variation between semantic and syntactic agreement. Chapter 7 is concerned with inter- and intra-speaker variation in spoken discourse, with variation related to register and genre, as well as with variation introduced by (re-)construal of the referent of noun and/or pronoun. Chapter 8 adds syntactic factors and explores the contribution of each factor to agreement choices made by a particular speaker under particular linguistic circumstances.

Chapter 7

Variation

The Individuation Hierarchy introduced in Chapter 6 can now be used to chart the patterns of variation in pronoun usage. Starting with intra- and inter-speaker variation in spoken language, then moving on to variation induced by different construal of the referent, the chapter also sketches semantic agreement in different genres of written language. We will close with some remarks on ambiguity arising from switched pronouns.

7.1 Inter- and intra-speaker variation

The spoken language data shows much variation between speakers, but variation in the speech of a single person is also not uncommon. This confirms the initial hypothesis that speakers indeed feel - to varying degrees of awareness - that there is a choice among pronouns. Moreover, intra-speaker variation suggests that this choice is made on the spot as the utterance is being formed. Apparent inconsistencies result, as the choice depends on a variety of factors.

First of all, each gender switch in itself constitutes a case of variation, since switching means divergence from the syntactic gender of the noun and often occurs side by side with non-switched attributives or other pronouns. Examples (1) illustrate the situation. In (1a), the referent is an animal, while b) concerns a (specific) mass referent. In both cases, the speaker uses first a syntactically agreeing pronoun and later switches to semantic agreement.

(1) a)   dat   **beest**   **dat**   moet je   niet gaan beklemmen
             DEM.N   animal(N) DEM.N must you not   go   hem_in

            want   dan   geeft **ie** een keer een   knauw.
            because then   gives 3.M one time a   bite

            'That animal, you can't tie him up, because then one day he'll snap'
            (CGN session 6789)

    b)   deze   week   hadden ze   verse **pasta**   in de
            this   week   had   they   fresh pasta(C) in   DEF.C

            aanbieding. en **die**   is gewoon keilekker. dus **die**
            sale(C)   and DEM.C is just   delicious   so   DEM.C

heb	ik	ook	gewoon	vier	pakken	meegenomen.
have	I	also	simply	four	packs	taken

't	is	toch	tot	eind	oktober	houdbaar
3.N	is	after_all	until	end	October	durable

'This week for example, they had pasta on special offer. And that's just delicious, so I took four packs. It'll keep until October anyway'
(CGN session 6996)

In both examples, the direction of the switch is as expected: (1a) contains a switch from neuter to masculine because the referent is an animal, while in (1b) it proceeds from common to neuter because the pronouns refer to a mass entity (albeit a specific mass). Yet, it is unclear so far why some pronouns agree syntactically while others agree semantically. As the individuation of the referent does not change within the individual utterances, semantics only predicts the direction, not the place of the switch. The examples in (1) show that speakers revise their choice in pronominalization at different points during sentence production. The issue will be addressed in Chapter 8.

Moving on from intra-speaker variation to inter-speaker variation, an example is given in (2).

(2)	dan	zie	je	zo	dat	**beeldscherm**.	en	dan	denk	je
	then	see	you	so	DEM.N	screen(N)	and	then	you	think

wow	wat	is	**dat**	klein. -	nee	ik	denk	eigenlijk
wow	what	is	DEM.N	small	no	I	think	actually

alleen	maar	bij	jouw	**computerscherm**
only	but	with	your	computer_screen(N)

wat	is	**ie**	groot.
what	is	3.M	big

'Look, then you see that screen and then you think, wow, this is really small. - No, actually, I only think, your computer screen, it's really big'
(CGN session 471)

Here, both speakers use (virtually) the same neuter gender noun but pronominalize it differently: the first picks a syntactically agreeing neuter gender demonstrative, the second opts for a semantically agreeing masculine personal pronoun. There is no real difference in individuation: the context suggests that both speakers have the computer screen in view and refer to the same entity.

This is a telling example because it defies the natural tendency for speakers to adjust their speech to one another or to be primed in various ways by their interlocutor. In

particular, the relevant clauses in the adjacent sentences are syntactically parallel, and within a context of syntactic priming (e.g. Bock 1986, Branigan, Pickering and Cleland 2000) the use of the neuter in the second utterance should be favoured.[1] Yet, the second speaker chooses a masculine pronoun because the referent is a bounded object.

In dialogue, where the pronominalization strategies of two speakers can affect one another, there are two possible scenarios. Where preferences differ, one speaker can adjust his/her speech to that of the interlocutor, or he/she can persist, and pronoun choice fluctuates between the speakers. In the corpus, there is evidence for both strategies.

For both intra- and inter-speaker variation it holds true that changes in pronoun choice usually proceed from syntactically motivated to semantically motivated. Switching in the other direction is rare. An example is (3). The first speaker starts with a cataphoric common gender pronoun, but the second speaker uses neuter gender pronouns that match the neuter gender noun syntactically.

(3) A: ik   ken   **deze**   helemaal   niet   dit   **boekje**.-
          I   know   DEM.C altogether not   DEM.N booklet(N)

     B: nee **'t**   is   ook   vrij   nieuw. **'t**   stond er   opeens.
          no   3.N is   also   quite new   3.N   stood there   suddenly

        'I don't know this at all, this booklet. - No, it's quite new, all of a sudden it stood there.'
        (CGN session 627)

Yet more unusual are reverted switches, i.e. cases where a speaker changes from syntactic to semantic agreement and back again.

(4) A: we   hebben daar zo'n   heel   **boek**   gemaakt van
          we   have   there such_a whole book(N) made   of

          onze   verkeringstijd.
          our   engagement_time

     B: maar **die**   heb ik niet 'k   weet niet waar   **het** is.
          but   DEM.C have I   not   I   know not   where 3.N   is

---

[1] I am not aware of research on agreement feature priming between speakers, and as features in agreement are usually not open to variation, this is not an obvious field for investigation. Dutch provides a unique testing ground for such research, which should be of interest to psycholinguists and morphologists.

137

'We made a whole book about the time of our engagement. - But I don't have it... I don't know where it is.'
(CGN session 294)

Such data is interesting for Chapter 8 where the constraints and preferences in switching behaviour are discussed.

Finally, there can even be switches in exophoric reference, i.e. in antecedentless anaphors. Here is an example. The referent is the device that is recording the conversation.

(5)   de   auto's neemt **ie**   ook op. en   als we naar jouw
       the   cars   takes 3.M also on and if   we to   your

       kamer   gaan dan neemt **het** ook auto's op.
       room(c) go   then take   3.N also cars   on

       'It's recording the cars, too. And when we go to your room then it also records the cars.'

Here, the referent does not change, but the anaphors fluctuate between masculine and neuter.

In spoken language, we encounter a multitude of variational patterns. This confirms that speakers feel the freedom to choose among several pronominal genders. Yet, pronoun choice and switching generally happen without reflection or awareness, and speakers are often surprised when their linguistic behaviour is pointed out to them. The fact that such massive variation attracts so little attention confirms the naturalness of the process and the wide spread of the phenomenon.

7.2 Gender switching in written language

The corpus provides insights into pronoun usage in spoken discourse, but gender switching does not stop at the gates of speech. While speakers use pronouns unselfconsciously in conversation, pronoun choice is often considered difficult in writing. Where dictionaries are not at hand or where the clash between expected norm and personal intuition is too severe, spoken language-type choices sneak into writing. This is most noticeable in informal texts such as e-mail, on websites, in letters to newspaper editors and, generally, in poorly edited print media. Another significant source are advertising texts, which often exploit the familiarity of colloquial style and thus contain pronouns that are closer to the spoken than to the written register. While a systematic study of written corpus data was outside the scope of the present research, this section presents some impressionistic evidence, sorted by text type.

7.2.1 Newspapers and books

Printed texts about a particular subject that can be expressed with a noun are excellent sources for varying pronoun usage. A fine example is a text about sugar consumption, found in a respectable magazine. In (6), we see the following pronominalizations of the common gender noun *suiker* 'sugar'.

(6) a) Deze **suiker** wordt vaak toegevoegd als zoetmaker aan
            DEM.C sugar(C) is   often added   as sweetener to

            tal   van voedingsmiddelen zoals   [...]. **Het** zit ook in
            a_number of   groceries   such_as   3.N sits also in

            minder voor   de hand liggende producten
            less   in_front_of the hand lying   products

            'Sugar is added to a number of foods as a sweetener, such as [...]. It is also contained in some less obvious products'

    b) **Suiker** levert   niets   gezonds. **Hij** is een belangrijke
            sugar(C) delivers nothing healthy   3.M is an   important

            energieleverancier en   daarmee kan **hij** overgewicht veroorzaken.
            energy_deliverer   and with_that can 3.M overweight   cause

            'Sugar gives you nothing healthy. It's an important energy provider, and so it can cause overweight'

    c) Ook voor **suiker** geldt de stelregel: zolang   u   **dit**
            Also for   sugar(C) holds the rule   as_long_as you 3.N

            met mate   binnenkrijgt...
            with measure intake

            'Also for sugar, the rule holds: as long as you consume it in moderation'

            (all examples from *Consumentengids* 1/2005)

Here, the common gender noun *suiker* 'sugar' is pronominalized as masculine in accordance with its historical gender (still given in the standard spelling dictionary, the *Woordenlijst*), but between masculine pronouns the neuter gender pronouns appear that colloquial language prefers for mass nouns. The really interesting point here is how well the distribution of the two genders lines up with semantics. In (6a and c), the sugar has a passive role in the described event: it is a stuff contained in some other stuff or in the body. In (6b), by contrast, the role of sugar is much more active, as a provider of energy and a causer of overweight. In this latter meaning, it is pronominalized as masculine.

Similar variation can be found with the noun *kaas* 'cheese'. Example (7a) and b) were found in the same (newspaper) article.

(7) a) Hoe langer de   **kaas**   mag rijpen, hoe harder
          how longer DEF.C cheese(C) may ripen   how harder

          die   wordt
          DEM.C becomes

          'The longer the cheese is allowed to mature, the harder it becomes'

     b) **kaas**   bevat   ook veel   goede eigenschappen.
          cheese(C) contains also many good   properties

          Zo is **het** rijk aan calcium
          so is 3.N   rich in   calcium

          'Cheese also has a lot of positive qualities. For example, it's rich in calcium'

          (both from newspaper *Sp!ts*, 26-01-2005)

Again, the influence of individuation is clearly visible. In (7a), the cheese is a sort or even an individual cheese, while b) refers to cheese in general. The pronominal genders neatly match the difference.

Newspapers and magazines also occasionally provide evidence for the inconsistent use of masculine and feminine gender for animals. Two examples are given in (8) and (9). In (8), the referent is a snake of unknown sex (and *slang* 'snake' used to be a feminine noun), while the pronouns in (9) refer to a female cat. The switch from feminine to masculine pronouns in the presence of another antecedent candidate, the common gender noun *zoon* 'son', leads to hilarious ambiguity.

(8) Hoe sterk het gif   is, hangt af van de   leeftijd van de
     how strong the venom is hangs off from the age   of   DEF.C

     slang,   wanneer **zij** voor het laatst heeft gegeten, hoe diep
     snake(C) when   3.F   for   the last   has   eaten   how deep

     de   giftanden   in het lichaam van **zijn**   slachtoffer dringen
     the venom_teeth in the body   of   POSS.M victim   penetrate

     en   de   hoeveelheid gif   die **hij** inspuit.
     and the amount   venom that 3.M   inject

140

'The strength of the venom depends on the age of the snake, when it has eaten last, how deep the venom teeth enter the body of its victim, and the amount of venom that it injects'
(*National Geographic*, Dutch/Belgian edition, May 2005)

(9) **Ze**   was de **poes** van onze zoon Bob, die   we nog gebeld
     3.F   was the cat(C) of   our   son Bob REL.C we PRT phoned

     hebben om   op tijd te komen voor **haar** sterven. Maar
     have   in_order_to in time to come for POSS.S dying   but

     hij moest uit Tilburg komen, dus hij kwam net een
     3.M had_to from Tilburg come   so 3.M came just an

     uur te   laat. We hebben **hem** met z'n allen begraven.
     hour too late we have   3.M with our all   buried

     'She was our son Bob's cat, whom we called so that he would arrive in time for her death. But he had to come all the way from Tilburg, and so he was an hour late. We buried her (lit.: him) together.'
     (magazine *Opzij*, cited in *Onze Taal*, 5/2005)

Probably the most common instantiation of semantic agreement in written language are non-neuter pronouns for neuter nouns in reference to persons. This is illustrated by the following newspaper headline from the large daily newspaper *Volkskrant*.

(10) Den Haag   geschokt over **slachtoffer** Schipholbrand   **die**
     The Hague shocked about victim(N)   Schiphol-fire(C) REL.C

     onterecht   vast zat
     wrongfully fixed sat

     'The Hague shocked about victim of Schiphol-fire who was detained wrongfully'
     (*Volkskrant*, 17-12-2005)

While examples from newspaper texts can often be blamed on poor editing, instances of mismatching pronouns can also be found in books and other carefully edited printed matter. There follow two examples from modern (children's) novels.

(11) De   **koffie**   is vies, maar Madelief drinkt **het** dapper op.
     DEF.C coffee(C) is nasty but Madelief drinks 3.N bravely up
     'The coffee tastes nasty, but Madelief drinks it bravely'
     (Guus Kuijer 1983 "Het grote boek van Madelief" Querido, p. 293)

(12) Ze   wreven zich   niet in met **azijn**,   maar dronken **het**.
they   rubbed themselves not in with vinegar(C) but   drank   3.N
'They didn't apply the vinegar to their skin but drank it'
(Thea Beckmann 1998 "Geef me de ruimte" Lemniscaat, p. 68)

Both cases exactly correspond to spoken language patterns.

Even in non-fictional literature, pronoun switches can be encountered. The following example comes from a book for post-graduate students. It is an interesting case because *verlenging* 'extension (here: extra contract time)' is a nominalization with the suffix *-ing*, allegedly associated with feminine gender. Yet, the pronouns are neuter (when extensions in general are referred to) and masculine (a specific extension).

(13) Dat is dan ook de reden   om nooit op een **verlenging**   te
that is then also the reason to   never on an   extension(C) to

rekenen. Zorg ervoor dat je   **het** niet nodig heb.
count   care for   that you 3.N not need have

Knijp   in je   handjes als je **hem** onverhoopt   krijgt.
squeeze in your hands   if you 3.M   unexpectedly get

'This is also the reason why you should never count on extra time. Make sure you do not need it. Enjoy it if you get it unexpectedly'
(Herman Lelieveldt 2001 "Promoveren" Amsterdam: Aksant)

Finally, switched pronouns can even occur in the genre characterized by the most carefully chosen wording: poetry.

(14) Als een **gedicht**   kan denken
if   a   poem(N) can think

denkt **het** dat ik kan doodvallen
thinks 3.N that I can drop_dead

wat   **hem** betreft.
what 3.M   concerns

'If a poem can think/ it thinks that I can drop dead/ as far as it's concerned'
(Toon Tellegen 2004 "Denken en kunnen" from "Minuscule oorlogen" p. 44)

This example is striking because this tiny part of the poem contains pronouns with two different genders, first a neuter (syntactically agreeing with *gedicht* 'poem'), then a masculine. It is especially interesting because in the third line, the neuter *het* is not available for other reasons: the pronoun slot comes with heavy stress, while

het is inherently unstressed. Yet, the construction is not avoided, but the masculine is pressed into service, just as would often happen in spoken language.

With regard to the above-mentioned sources, the impression arises that switches to neuter gender pass editors' eyes more easily than switches to masculine or common gender. Speakers seem to regard them as less transgressive. Confronted with sentences such as (11) or (12) above, they often fail to see 'what is wrong'. When prompted to replace the neuter pronoun by another they are often unable to find an alternative. This suggests that acceptance is higher for switches to the neuter, while switches to masculine or feminine are more likely to attract attention and resistance.

7.2.2 Other public communication

A good source for pronoun variation are public texts such as advertisements, noticeboards and similar means of communicating with a large readership. The wording is usually very carefully chosen, but it can be in the interest of the author or the company to divert from the written standard in favour of a more colloquial style. Four examples follow, two for bounded referents, two with reference to masses.

(15) a) **Hij** ziet er inderdaad onschuldig uit. Dit **HiFi-systeem**
3.M looks PRT indeed innocent out DEM.N hi-fi_system

toont **z'n** ware karakter wanneer je **'m** aanzet.
shows POSS.M/N true character when you 3.M switch_on

'Indeed, it's looking innocent. This hi-fi system shows its true character when you switch it on.'
(billboard *Sony*, 2003)

b) Het decoratieve **bamboeplantje** geeft het huis een
def.n decorative bamboo_plant.dim(n) gives def.n house(n) a

vrolijke, zomerse uitstraling. Geef **hem** veel water!
happy summery atmosphere give 3.m much water

'The decorative bamboo plant gives your house a happy summery air. Give it a lot of water!'
(supermarket magazine *AllerHande* 08/2004)

(16) a) Als u **kleding** hebt gepast, wilt u **het** dan aub
if you clothing(C) have fit would you 3.N then please

naar de kassa terugbrengen?
to the checkout return

'If you have tried on clothing, would you please return it to the checkout?'
(note in a store)

b) biedt groene, ofwel zachte **zeep** dé oplossing voor moeilijke
offers green.C alias soft.C soap(C) the solution for difficult

vlekken. Bovendien is **het** een uitstekende ontvetter, laat
stains moreover is 3.N a great de-greaser lets

het tegels weer glanzen én helpt **het** tegen bladluis.
3.N tiles again shine and helps 3.N against plant_louse

'... green or soft soap is THE solution for difficult stains. Also, it removes grease, it gives tiles back their shine and it works against plant lice'
(supermarket magazine *AllerHande* 11/2006)

Again, we see the patterns of spoken language replicated: neuter gender count nouns appear with masculine or common gender pronouns, while common mass nouns get neuter *het* or *dat*.

7.2.3 Internet texts

The internet with its user-generated content is an open playground for variation. As a corpus, it is of little use, since the metadata is missing, in particular the background of the writer. Texts by native speakers alternate freely with those written by learners or translated by software, and among proficient language users, regional variation adds to the range. Given these facts, it is not surprising that virtually any combination of noun and pronoun can be found. The possibility to distinguish between erratic and systematic pronoun usage is the major advantage of research with the help of properly assembled and annotated corpora.

Yet, internet documents do provide interesting information. First, the sheer amount of available data allows checking for rare patterns, such as the pronominalization of uncountable abstracts or common-to-neuter switches on the relative pronoun. Thus, the following interesting cases were found.

(17) a) Constructief wordt **ironie** als **het** ingebed is in empathie
constructive becomes irony(C) if 3.N embedded is in empathy

'Irony becomes constructive when it is coached in empathy'
(www.wapenveldonline.nl/viewArt.php?art=545)

b) Saaie lezingen, donkere ruimten, zware maaltijden of lange
boring lectures dark rooms heavy meals or long

autoritten veroorzaken geen **slaperigheid**, ze   brengen **het**
car_journeys cause   not sleepiness(C) they bring   3.N

slechts naar de oppervlakte.
only   to   the surface

'Boring lectures, dark rooms, heavy meals or long car journeys don't
cause sleepiness, they just bring it to the surface'
(www.hartenziel.nl/.../diep_onder_zeil_en_toch_liggen_piekeren)

(18) a) Een motor werkt op brandende **benzine dat**   uitzet
          an engine works on burning   fuel(C) REL.N expands
          'an engine works with burning fuel that expands'
          (www.dbmengineering.nl/dbm/tuning/nx/index.htm)

      b) **rode lava   dat**   uit de   vulkaan   naar beneden
          red.C lava(C) REL.N from DEF.C   vulcano.C to   down

          stroomde
          streamed

          'red lava that streamed down from the vulcano'
          (http://noorski.waarbenjij.nu/index?page=message&id=1551644)

Moreover, the vastness of the data and the convenient search options help to identify collocations that need to be set aside for their lack of choice or failure to exhibit agreement. Examples are *op z'n kop* 'upside down', *op z'n beurt* 'in turn' or *in z'n geheel* 'as a whole', which all contain a masculine/neuter singular possessive, but combine readily with any sort of possessor noun, even plural nouns or nouns such as *kathedraal* 'cathedral' or *architectuur* 'architecture' that are usually treated as feminine in writing.

Third, online searches allow the comparison of particular combinations for which a corpus would not yield a sufficient number of examples. For example, Audring (2006a) reports an investigation on the influence of familiarity or kinship on agreement choices. To this aim, two internet searches were conducted, one with Google (www.google.com), the other with Webcorp (www.webcorp.org.uk). The search string was a neuter gender noun immediately followed by a relative pronoun.

Here, the test is repeated with a different setup. In order to introduce a familiarity divide, neuter nouns with person reference were searched for, combined with the attribute *dat* [N] 'that' for the non-familiar condition and *mijn* 'my' for the familiar condition, followed by pronominal *dat* [N] or *die* [C], which can be the relative pronoun 'who' or the demonstrative anaphor 'he/she'. *Dat* represents semantic agreement, while *die* agrees semantically. Frequencies were compared for both conditions. The search strings were *dat zusje dat*, *dat zusje die*, *mijn zusje dat* and *mijn zusje die* ('that/my sister who'). The nouns were *dochtertje* 'daughter.DIM',

zoontje 'son.DIM', *broertje* 'brother.DIM', *zusje* 'sister.DIM', *vriendje* 'boyfriend.DIM' and *vriendinnetje* 'girlfriend.DIM'.

The results are presented in chart (19). The columns show the percentage of semantic agreement as opposed to syntactic agreement for non-familiar (white) and familiar (black).

(19)
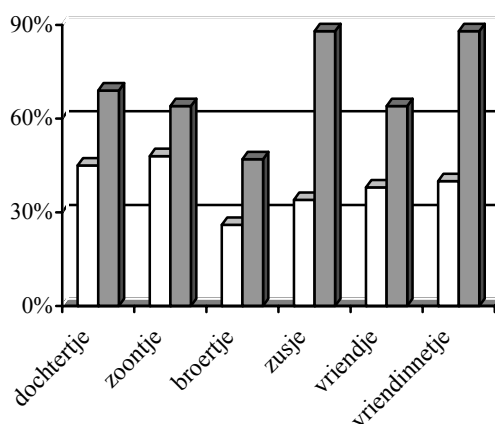
For all of the six nouns, the likelihood of semantic agreement is higher in the familiar than in the non-familiar condition. This suggests that familiarity or kinship promote the referent on the hierarchy and make a neuter pronoun a less attractive option.

Also, the differences between the agreement patterns of individual words can be investigated with the help of the large database that is the world wide web.

7.3 Variation as a consequence of construal

Looking back at the examples so far, we have seen a lot of alternation between semantic and syntactic agreement. The relative frequency of the two types of agreement will be discussed in the next chapter. First, the issue of ***construal*** will be addressed, which is the key to much of the variation observed.

In (6) and (7) above, we saw that a single noun can refer to entities with a different degree of individuation. This ties in with the well-known fact that countability is not a fixed lexical property of a noun but rather a characteristic of a noun phrase (Allan 1980) and open to different construals. In a count-mass-sensitive agreement system, this is an obvious source of variation.

Different construals can occur in many semantic fields. A few shall be mentioned in the following. Probably the most common case is generic-specific construal. An illustrative example comes from the Flemish part of the Corpus of Spoken Dutch.

(20)   dat   ze   **suiker**   mee   hadden   omdat   **dat**   zo   duur
       that   they sugar.C   with   had   because DEM.N   so   expensive

       kostte   in   Rusland
       cost   in   Russia

       'that they had sugar with them because it was so expensive in Russia'
       (CGN session 755)

In this utterance, the noun *suiker* 'sugar' refers to a particular, limited amount of sugar that was taken to Russia. The anaphoric pronoun *dat* 'that', which should be coreferent with the noun, refers to sugar in general. Both referents are covered by the semantics of *suiker* 'sugar', but the generically referring pronoun represents a lower degree of individuation than the noun.

Another common case is count-mass-related construal. This mostly involves individuating a mass referent into a portion, a sort or a container as in *my coffee*, *this tea* or *put the wine in the fridge*. In more exotic cases, word meaning can be extended metaphorically or metonymically, with accompanying change in countability. Consider (21).

(21)   je   hoeft   je   niet   verplicht   te   voelen   om   opeens
       you   need   you   not   obliged   to   feel   to   suddenly

       die   **vlierbloesemgelei**   te   maken.   -
       DEM.C   elderflower_jelly(C)   to   make

       denk   dat   **ie**   niet   in   het   kookboek   maar   in   ...
       think   that   3.M   not   in   the   cookbook   but   in

       'You don't have to feel obliged to make that elderflower jelly. I don't think
       it's in the cookbook but in...'
       (CGN session 602)

Here the pronoun refers not to the substance denoted by the noun, but to a recipe, which is a bounded abstract entity. This reading metonymically extends the semantic scope of the word *vlierbloesemgelei* 'elderflower jelly'. Yet the substance reading and the recipe reading are semantically close enough to allow an anaphoric relation between the noun and the pronoun.

Another context rife with construal ambiguities is that of games, where figures stand for persons and cards represent places or materials. The corpus contains an abundance of examples, as many conversations were recorded during game-playing

at home, in particular the popular bord game *Settlers of Catan*, where cards figure as particular raw materials.

(22) ik   wil   **graan**.   - nee **die**   heb ik   ook   nodig
      I   want   wheat(N)   no, DEM.C have I   also   need
      'I want wheat.' - 'No, I need that myself.'
      (CGN session 422)

Without this knowledge, there is no explanation for why the second speaker picks a common gender pronoun for the neuter gender mass noun *graan* 'wheat, grain'. The noun is ambiguous in referring either to the card or to the material, but the mass reading is not expected to trigger common gender pronouns.

Next, a referent can be construed as an object or as an event. (23) is a case in point.

(23) we   zouden   naar   The Bone Collector   gaan.   maar   **die**   was
      we   should   to   The Bone Collector   go   but   DEM.C was

      uitverkocht.- en   waar   ben je   dan   nu   heen   geweest.-
      sold_out   and   where are you   then now to   been

      The Cider House Rules.- en?- ik   vond   **'t**   wel   leuk.   **dat**
      The Cider House Rules   and   I   found   3.N   quite nice   DEM.N

      is niet   echt   spannend maar **'t**   is   wel   leuk.   maar
      is not   really   exciting but   3.N   is   quite nice   but

      Joost vond   **'t**   dus   saai. [...] nee **'t**   was   wel   leuk.
      Joost found 3.N   thus   boring   no   3.N   was   quite nice

      want   **die**   had   zeven Oscarnominaties   gehad
      because DEM.C had   seven Oscar_nominations had

      'Then in the end he did come with us because we were going to "The Bone Collector". But that was sold out. - and then were did you go to? - "The Cider House Rules" - And? - Well, I liked it. It isn't very exciting, but it's nice. But Joost was bored. No, it was okay. Because, that one 's had seven Oscar nominations'
      (CGN session 476)

In this example, a film is discussed as an event, and referred to by neuter pronouns. Yet, in the beginning and the end of the dialogue fragment, the film is more of an object, and the speakers revert to common gender pronouns. This confirms that events and propositions are the domain of the neuter, while common gender is employed for object reference.

In yet more quirky cases, construal can turn an object into an action. In the next example, the noun *servieskast* 'cupboard' is used metonymically in the context of moving house: the speakers are planning who is doing what. The non-literal meaning is 'unwrapping, putting up and stocking the cupboard'. This triggers the use of a neuter pronoun which is associated with events and propositions.

(24) de   **servieskast da**   's   dus   echt   iets   van   jou
      DEF.C   cupboard   DEM.N   is   thus   really   something   of   you
      'The cupboard, that's really your thing'
      (CGN session 251)

If the pronoun referred to the cupboard directly, the neuter pronoun would be unexpected since *servicekast* is a common gender noun referring to an object and should not switch towards the neuter.

Finally, construal can take the shape of personification, when inanimate objects are construed as animates and thus promoted on the hierarchy. The subcorpus contains a fine example.

(25) 't   was   een   intelligent   onderwijssysteem   dus   't   was
      def.n   was   an   intelligent   education_system   so   def.n   was

      een   **systeem   dat**   zinnetjes   terug   zei[...]   en   als
      a   system(n)   rel.n   sentences.dim   back   said   and   if

      je   dan   een   fout   maakte   zei   **ie**   bijvoorbeeld
      you   then   a   mistake   made   said   3.m   for_example

      nee   de   druk   moet   in   pascal   niet   in   atmosfeer.
      no   the   pressure   must   in   pascal   not   in   atmosphere

      'It was an intelligent education system, so it was a system that talked back to you, and if you made a mistake then it said, for example, no, pressure is measured in pascal, not in atmosphere'
      (CGN session 629)

This example, as all the others, presents difficulties in semantic classification, as the referents fit more than one semantic class. Yet, for most of the cases an explanation in terms of construal accounts for the variation, particularly between speakers and in larger stretches of utterance.

For the present study, it is especially interesting to note that construal is reversible. Thus, while pronouns usually change from syntactically to semantically agreeing rather than vice versa, differences in construal can trigger pronominalization with otherwise unusual patterns. When an entity is now construed as an object, then as a substance, pronouns are expected to oscillate accordingly. In such cases, a semantics-based option can trigger the gender that is also the lexcial gender of the

noun. This causes overlap between semantic and syntactic agreement, which makes it hard to distinguish whether a pronoun has been chosen on the grounds of syntax or of semantics.

7.4 Ambiguity and problems in anaphor resolution

Some switches are truly surprising. This is when the switched pronoun has the syntactic gender of another potential antecedent, thus creating a problem for anaphor resolution. In such a situation, speakers would be expected to refrain from switching in anticipation of communicative problems. While such a claim cannot be investigated systematically, there are examples where speakers choose a gender that is plainly misleading in terms of feature computation. Consider the following utterance, from personal observation (thanks to my sister-in-law for the pronoun, and sorry to my mother-in-law for the broken plate).

(26) toen heb ik de   **suikerpot**   op het   **bord**   gezet,
      then have I   DEF.C sugar_pot(C)   on DEF.N plate(N) put

      en   toen brak **die**
      and then broke DEM.C

      'then I put the sugar pot on the plate and then it broke'

In this utterance, the anaphor seems to agree in gender with *suikerpot* 'sugar pot'. Yet, the broken object was the plate. Apparently, the danger of ambiguity was felt less strongly than the motivation to use a common gender pronoun when speaking of an object.

One might object that spontaneous speech involves too little monitoring to detect and avoid such problems, but similar cases even appear in writing. Two instances are given in (27) and (28).

(27) Hoe staat het nu   met de **aandacht** voor **autisme**?
      how stands 3.N now with the attention(C) for   autism(N)

      Dat   is gigantisch   verbeterd.
      DEM.N is gigantically improved

      'How about attention for autism? That has improved gigantically.'
      (magazine of University of Leiden, *Mare*, May 2006)

(28) het   galaconcert   waarmee   het   **orkest**   in
      DEF.N gala_concert(N)   with_which   DEF.N orchestra(N) in

      aanwezigheid   van   de   **koningin haar**   honderdjarige
      presence(C)   of   DEF.C queen(C) POSS.F hundred_years

bestaan   vierde
existence(N)   celebrated

'the gala concert when the orchestra was celebrating its (lit.: her) hundredth
anniversary in the presence of the Queen'
(Haagsche Courant 26-11-2004)

In (27), *aandacht* 'attention' is an unbounded abstract with common gender. Its
semantics suggests neuter anaphoric gender. Unfortunately, the intervening noun
autisme 'autism', has neuter syntactic gender, so the pronoun seems to refer to this
noun (which is also closer and therefore a more likely antecedent). Here, the writer
has not noticed the ambiguity or has opted for this risk rather than choosing a
syntactically agreeing, but semantically ill-fitting pronoun. The line in (28), which
native speakers find funny, contains an instance of the possessive pronoun *haar*
'her' as is popular in journalistic writing (see section 3.1.6 above). Here, it refers to
the neuter noun *orkest* 'orchestra', which is an infelicitous combination, especially
in a context with a much better antecedent for a feminine pronoun, viz. *koningin*
'queen'. While the choice is triggered by different factors than those discussed in the
preceding chapters, the consequence is the same: pronominalization is felt to be an
issue, and language users end up choosing pronouns that are ambiguous and cause
communicative difficulties. Such cases again illustrate the severity of the Dutch
pronominalization problem, and the need for a solution.

7.5 Conclusion

This chapter focuses on one of the central traits of the Dutch pronoun genders:
variation. Speakers vary not only amongst each other, but also individually between
one utterance and the next. The only clear limit to variation is that once syntactic
agreement has been abandoned for the sake of semantic agreement, switching back
is unusual.

Gender switching is also shown to occur in written text, not only on the internet,
where all sorts of linguistic behaviour can be expected, but also in print, from novels
to billboards and from product labels to poems. Sometimes, spoken-language type
pronouns are exploited for advertising purposes. Other cases are more accidental.
Especially neuter pronouns for mass nouns seem to be tolerated more or noticed less
than other types of switch.

It is argued that much of the variation reflects different construals of the nominal
referents. Fish is pronominalized differently when it is an animal than when it is an
ingredient of a meal, films can be construed as objects or events, and machines can
be promoted to near-human status. All of this can be reflected in pronoun choice.

The chapter ends with a brief discussion of switching and ambiguity.

Chapter 8

System Competition

The preceding chapters concentrated on switched pronouns, i.e. on pronouns that do not have the same gender as their antecedents. With the help of corpus data, patterns in switching behaviour were identified that answered the question which alternative gender is chosen by speakers and why. We now turn to the question of when, and under which circumstances, speakers opt for a switched rather than a matching pronoun. This chapter will also provide data on the relative frequency for syntactic and semantic agreement.

The Dutch situation can be seen as a competition between two systems. The first is the syntactic gender agreement system characterized by a feature match between a nominal antecedent and an anaphoric pronoun. Agreement is triggered by a morphosyntactic property of the noun, its lexical gender, which is faithfully mirrored in the gender of the agreement targets. This system works independently of semantics. While meaning may play a role in the process of assigning a noun to a gender, it does not influence the computation of agreement in speech production. The syntactic system has two gender values: common and neuter. The second system, by contrast, is semantic in nature, and agreement is based on the conceptual properties of the referent. The semantic system has four gender values: common, neuter, masculine and feminine, each with their own distributional pattern (Chapters 5 and 6). Distribution is largely independent of the lexical gender of the noun. The two systems correspond to what is known as syntactic versus semantic agreement.

In many cases, both systems lead to the same gender choice. When a noun has common gender and denotes a countable entity, syntactic agreement will favour a common gender pronoun, as will semantic agreement. The same holds for neuter gender mass nouns, which in any case get neuter gender agreement. Under such circumstances, we cannot say which of the two systems is responsible for the pronoun choice. Yet, there are a fair number of instances where semantic and syntactic gender agreement produce a different outcome. This situation arises with neuter gender count nouns and with common gender mass nouns. For such nouns, the two systems can be seen in competition (Table 1).

(1) Conflicting gender choices

	Gender of pronoun according to syntactic agreement	Gender of pronoun according to semantic agreement
Neuter gender count noun	*neuter*	*common, masculine*
Common gender mass noun	*common*	*neuter*

The goal of this chapter is to investigate under which circumstances one system is preferred to the other and which factors are relevant for the choice.

From the subcorpus of 500,000 words, all pronouns were collected, whether semantically or syntactically agreeing, and a number of criteria was tested for their potential influence on the choice between one switch (semantic agreement) and match (syntactic agreement). After a brief discussion of the relevant criteria and the expectation of their influence, the contribution of each is evaluated.

8.1 Factors influencing pronoun choice

There are several factors that are candidates for exerting influence on the pronoun choice in a particular situation. They can be roughly sorted into four categories. The first has to do with the pronoun, the second with the distance between pronoun and antecedent noun, the third concerns the antecedent itself and the fourth the speaker. We will discuss each candidate factor in turn.

8.1.1 The pronoun

In the corpus study, three different sorts of pronoun were investigated: personal pronouns (including demonstrative pronouns in anaphoric use), relative pronouns and possessive pronouns. Especially the first two types are interesting because each occupies a different agreement domain, as expressed on the Agreement Hierarchy (the possessive does not feature on the hierarchy). According to the Agreement Hierarchy, semantic agreement is expected to be more frequent with personal than with relative pronouns. In 8.3 below, it will be shown that the Dutch data confirms the predictions.

For the possessive pronouns, the research is hampered by their extremely low frequency in reference to inanimate entities. Also, their impoverished morphology allows them to distinguish only between feminine and non-feminine gender, which means that they cannot be used as evidence for the influence of countability, as this split is expressed in the choice of neuter gender on the one hand and masculine or common gender on the other. Yet, there is an interesting usage fact that pleads for some attention to possessive pronouns. This is the fact that in person reference, semantic agreement is considered the only possible and 'correct' option. If this were generally true, possessive pronouns would be the most progressive agreement target of all. Section 8.3 below shows that this pattern is corroborated by the corpus data, although the numbers are too low to allow generalizations.

Demonstrative pronouns in anaphoric use occupy the same syntactic slot as personal pronouns. Therefore, both types were considered together. Yet, it is not unlikely that different forms of pronouns exhibit different agreement behaviour. Therefore, the switch/match ratio was measured for three types of anaphoric pronoun: full form pronouns (*hij, hem, zij, haar, het*), clitic or reduced pronouns (*ie, 'm, ze, 'r/d'r, 't*) and demonstrative pronouns (*die, dat, deze, dit*). Intuitively, full form pronouns, being more strongly associated with formal discourse, are expected to be more

conservative than clitic/reduced pronouns or demonstrative pronouns. Figures are given in section 8.3.

Last but not least, it has been noted in the relevant literature that the preference for semantic or syntactic agreement may differ for subject pronouns and object pronouns. As Corbett puts it:

> Whenever, in a given position on the Agreement Hierarchy, there is a difference between the agreements found in the nominative and in the oblique cases, the likelihood of semantic agreement in the nominative will be as high as or higher than the likelihood of semantic agreement in the oblique cases.
> (Corbett 1991: 238)

There is interesting data from pronoun usage in English dialects (Ihalainen 1985, 1991, Wagner 2003, Siemund 2008). Ihalainen and Siemund interpret the differences in agreement behaviour to semantics rather than to morphosyntax by attributing them to the Accessibility Hierarchy (Keenan and Comrie 1977) which captures the typological fact that subjects are generally higher in salience, animacy, individuation and/or definiteness (the discussion is complicated by the partial overlap between the notions and terms). According to these considerations, semantic agreement should be more likely for count nouns in subject position than in object position. For mass nouns, the prediction is less clear. Their overall propensity to agree semantically should be less because they are inherently less individuated, but they might switch more when they are subjects. The effect of subjecthood and objecthood on switching was tested for anaphoric demonstrative and personal pronouns together. The test was then repeated for both sorts of pronoun in isolation in order to see if the expected effect is indeed due to subject or object status, or if there is an influence of case marking (nominative versus oblique). In Dutch, only the personal pronouns mark case morphologically; the demonstratives do not. The data is discussed in 8.3 below.

8.1.2 Distance

One factor that is well known to have an influence on the choices between syntactic versus semantic agreement is the distance between noun and pronoun (although the literature contains few empirical studies, one of the exceptions is Levin 2001). Corbett (1979: 220) finds that 'real' distance alone is not sufficient to predict the agreement behaviour of different agreeing elements, but that distance effects have predictive value within a single type of agreement target (say, personal pronouns). Thus, distance measures were compared for the anaphoric pronouns (i.e. personal and demonstrative pronouns). As explained in Chapter 4, two distance measures were used, a rough structural measure and a more fine-grained word count. Structural distance was coded in four values: *within clause*, *within sentence*, *within turn* and *beyond*. Word count captures the number of lexical items intervening between the antecedent noun and its pronoun, the minimal distance being zero (adjacency).

It is predicted that greater distance according to both measures correlates with higher likelihood of semantic agreement. This effect is expected to hold on sentence as well as on corpus level. This means that in an utterance with more than one personal pronoun, semantic agreement should increase in likelihood from the pronouns that are closest to the antecedent to those that are further removed. This also means that switches generally proceed from syntactic to semantic agreement rather than vice versa. Individual cases contradicting this pattern should be in the minority. On the level of the corpus, the overall distance between pronoun and antecedent is expected to be smaller for syntactically agreeing than for semantically agreeing pronouns. This can be calculated as a general mean for all records in the database. An alternative is to separate the data into two parts, one consisting of those records containing an antecedent and a single agreeing pronoun, the other comprising all cases where an antecedent is followed by several pronouns. This separation shows the effect of distance with or without the influence of inter-pronoun effects that might skew the picture. In both cases, semantically agreeing pronouns are expected to be further removed from their antecedent than syntactically agreeing pronouns.

Another issue that may be considered is the influence of word order. When the distance is negative, i.e. when an agreement target precedes its controller, syntactic agreement is predicted to be more likely than for an equivalent postnominal target (Corbett 1979: 217 ff). Unfortunately, there are only 16 cases of preposed pronouns (cataphors), only two of which are matches, so the data is insufficient for statistical analysis.

The tables in section 8.3 give the figures relating to distance effects, as well as an overview of the switching patterns (syntactic to semantic, semantic to syntactic).

8.1.3 The noun

While the noun is, of course, the predictor for which pronominal gender is chosen, it may also have an influence on **when** one option or another is preferred. Thus, the nouns *mannetje* 'little man' and *boekje* 'booklet' have the same agreement options, viz. neuter (because they are neuter nouns; syntactic agreement), common and masculine (because they denote countable entities; semantic agreement), but they may differ in how often the choice is made for one option or the other. Two properties of the noun come to mind as worthy of investigation.

The first property concerns the semantics of the noun. Chapter 6 argues that the distribution of Dutch pronouns is sensitive to the *degree of individuation* of the NP's referent. A high degree of individuation is associated with feminine, masculine or common gender, while a low degree correlates with neuter gender. These semantic associations can cause speakers to feel that a noun has the wrong gender for the purposes of pronominalization. Conflicts between referent semantics and noun gender are expected to be strongest at the extreme ends of the individuation hierarchy, i.e. for persons on the left end and for unspecific masses on the right. If this is so, we expect more instances of semantic agreement for nouns on the extreme left and the extreme right of the scale, while those in the middle should be perceived

as less problematic and therefore fluctuate more between the two agreement sorts. This prediction was tested, and the results are given in 8.3.

The second property concerns the morphological structure of the noun. In many cases where a neuter gender noun takes a non-neuter pronoun, the antecedent is a diminutive noun. As sketched briefly at the end of Chapter 4, the gender of the root noun might be active in the speaker's mind, which then can infuence the agreements. In order to isolate this factor, diminutive nouns were considered separately. Table (15) below compares the number of switches for diminutives with neuter gender roots to those with common gender roots. If the gender of the root is influential, nouns with common gender roots should be more likely to switch to common or masculine gender.

8.1.4 Sociolinguistic factors: age of speaker

With the general restriction to spontaneous speech, most sociolinguistic factors, particularly register and style, have been controlled for. While they are undoubtedly of tremendous influence, they exceed the scope of the present study. Yet, there is a sociolinguistic factor that is easily available through the corpus data and promises interesting evidence. This factor is the age of the speaker. Mapping linguistic behaviour across several age groups is, of course, a well-known technique to tap into the diachrony of recent language change. This *apparent-time* method is also useful here. Some pronominalization strategies, notably the masculine pronoun for neuter nouns, are felt to be a relatively recent phenomenon, so they are expected to be more frequent in younger speakers. In order to see if pronominalization choices differ between the generations, all speakers were sorted into six age groups: below 20 years of age, between 21 and 30, 31-40, 41-50, 51-60 and above 60. If the prediction is correct, the switch/match ratio should change gradually from the lowest to the highest age group, with the relative number of switches monotonically decreasing. Table (16) in 8.3 gives the results.

8.2 Two data sets

Most of the results of the corpus study are given for two conditions on the data set. This has to do with the treatment of the common gender count nouns. In the original setup of the research, only those instances were counted where the speaker had a choice between one gender or another. This includes those cases listed in Table (1) above, and excludes neuter gender mass nouns and common gender count nouns which, after all, do not trigger any switching pronouns.

A difficulty is that common gender count nouns do offer a choice, but only between common and masculine gender. Pairs of common gender nouns and masculine gender pronouns are uncertain in their status as switches. In Chapter 5, it was argued that the distribution of masculine pronouns with common gender antecedents is best described in semantic terms. Yet, there are reasons why common-to-masculine switches are not exactly equivalent to neuter-to-masculine or common-to-neuter switches.

Theoretically speaking, a switch from common gender to masculine gender does not contradict syntax in quite such a drastic way as a switch from common to neuter or from neuter to common gender. This holds especially when common gender is seen as a fusion of masculine and feminine.

On semantic terms, common gender count nouns are not conflictive in the sense that they have the 'wrong gender' for their semantics. Conflictiveness is important for the present study because for those nouns that lack it, pronoun choice on semantic grounds overlaps with those on syntactic grounds. Thus (2a-b) could have been produced by syntax as well as by semantics.

(2) a) de   **vriezer**   maakt   een   hoop   lawaai   hè? -   zou
         DEF.C   freezer(C)   makes   a   lot   noise   he   should

         die   nou   ook   's   nachts   zoveel   lawaai   maken? -
         DEM.C   now   also   at   night   so_much   noise   make

         hij   is   nu   al   een   hele   tijd   niet   open   geweest
         3.M   is   now   already   a   whole   time   not   open   been

         'The freezer makes a lot of noise, doesn't it? - Do you think it's making so much noise at night, too? - It hasn't been open for quite a while now.'
         (CGN session 309)

     b) het   **water**,   **dat**   is   ook   met   zwavel
         DEF.N   water(N)   DEM.N   is   also   with   sulphur
         'the water, that's with sulphur, too'
         (CGN session 262)

As common and masculine pronouns are covered by the same rule - both are associated with animate and/or countable entities - they also violate the conflict condition. Yet, for the same reason, common and masculine pronouns often appear together in a single agreement chain. (2a) is a case in point. Thus, excluding common gender pronouns but including masculines is not an option.

There are two more problems. First, as pointed out earlier, the masculine nominative clitic pronoun *ie* is formally indistinguishable from the common gender demonstrative *die* when the preceding lexical element ends in a /t/. As anaphors to common gender nouns, the masculine is regarded as a switch, the common gender pronoun as a match. Being formally indistinguishable, such cases are inconclusive. A second difficulty with masculine pronouns is that there is no masculine gender in the relative pronouns. Thus, counting masculine pronouns with common gender antecedents as switches rather than matches boosts the number of personal pronoun switches without a chance for the relative pronouns to catch up. This is because there is only one way in which a relative pronoun with a common gender antecedent can be a switch: by having neuter gender. Personal pronouns, by contrast, can be switches when they are masculine, feminine or neuter.

158

Excluding common-to-masculine switches solves both problems. Yet, as the theoretical choices are difficult and fallible, it was chosen to present results, whenever relevant, with two conditions on the data set: once with the full set and once with common-to-masculine switches excluded. For the sake of consistency, the full data set also entails all nouns that do not cause conflicts between gender and meaning and therefore never switch. The two conditions will be referred to as "restricted data set" and "full data set". The following table summarizes the differences.

(3) Conditions on the data set

Noun	Restricted data set	Full data set
Common count	*excluded*	included
Common mass	included	included
Neuter count	included	included
Neuter mass	*excluded*	included

The restricted data set contains 810 records by 80 different speakers, the full set 1630 records by 89 speakers.

Note again that common gender human nouns such as *moeder* 'mother' were excluded across the board because of their immense frequency as antecedents in pronominalization and their lack of variability.

8.3 Results of the corpus study

The following section presents the results of the corpus study. Figures are given in the form of contingency tables, with number of switches (semantic agreement) and matches (syntactic agreement) compared for each of the factors listed in the previous section.

8.3.1 The pronoun

First among the pronoun-related factors is the influence of the sort of pronoun on the frequency of switching. Two pronoun sorts are compared: relative and personal pronouns (including demonstrative anaphors). Agreement behaviour was tested for both data sets. The Agreement Hierarchy predicts that relative pronouns are less likely to switch than personal pronouns.

(4) Switching according to the AH

a) Restricted data set

	Relative pronoun	Personal pronoun (including anaphoric demonstrative)
Switch	21 (25%)	469 (65%)
Match	62 (75%)	251 (35%)

$$\chi^2(1) = 49.66 \ p<.001, \ \text{Cramér's V}=.25$$

b) Full data set

	Relative pronoun	Personal pronoun (including anaphoric demonstrative)
Switch	21 (13%)	871 (60%)
Match	143 (87%)	588 (40%)

$$\chi^2(1) = 130.97 \ p<.001, \ \text{Cramér's V}=.28$$

The data shows that the prediction is borne out in both conditions. The relative pronouns switch - at most - a quarter of the time, while personal pronouns and anaphoric demonstratives switch at least sixty percent of the time. The results are significant for both conditions. This is shown by the results of Pearson's χ^2-test. The value p is smaller than 0.001, which means that it is very unlikely that the correlation between pronoun type and switching behaviour is a coincidence. A second calculation, Cramér's V, was executed in order to evaluate the strength of the correlation. Here, a value of 0.28 indicates an effect of medium strength (values closer to 0 are interpreted as weaker, values closer to 1 as stronger effects; interpretation according to Field 2005). This means that the sort of pronoun is a medium good predictor for switching behaviour. In other words, *het boek... die* 'the.N book(N)... that.C' is almost three times as likely if *die* 'that' is an anaphoric demonstrative rather than a relative pronoun, and this is not due to chance, but to a systematic link between pronoun sort and gender preferences.

The difference is more marked in the full data set, i.e. when masculine pronouns following common gender nouns are counted as switches and when all matches, even the unconflicting ones, are considered. The reason is that extending the data set does not increase the number of relative pronoun switches along with that of relative pronoun matches because there are no masculine gender relative pronouns. The problem was addressed above. For the personal pronouns, the switch/match ratio is roughly the same in both conditions.

Next, the agreement behaviour of the possessive pronoun was compared to that of the other pronouns. This test was carried out for a subset of the data, viz. neuter nouns referring to female persons. The reason is that possessive pronouns do not distinguish masculine and neuter gender, so possessive pronoun switches can only manifest themselves for neuter nouns with female referents. It is expected that the likelihood of switches increases from left to right in the table.

(5) The behaviour of the possessive pronoun

	Relative pronoun	Personal pronoun + demonstrative	Possessive pronoun
Switch	2 (25%)	61 (97%)	7 (100%)
Match	6 (65%)	2 (3%)	-

$$\chi^2(2) = 40.67 \text{ } p<.001, \text{ Cramér's V=.72}$$

The prediction is borne out and the χ^2-test indicates significance, but the number of instances is too low for the relative pronouns and for the possessives. The problem here is the **expected count**, which is too small. Expected count is a calculated value that expresses for each cell in the table how many instances one would expect if the null-hypothesis were true: correlations within the data are mere chance. If the expected count is < 1, the calculation is invalid. Thus, the issue needs to be investigated with more data than was available for the present study.

Next in line, the question was asked whether the form of the pronoun could have an influence on its switching behaviour. The hypothesis was formulated that full form pronouns may be more conservative and thus more reluctant to switch, while clitic pronouns and demonstrative pronouns may be more progressive. As before, the issue was tested for both data sets.

(6) Preferences depending on form of pronoun

a) Restricted data set

	full pronoun	clitic pronoun	demonstrative pronoun
Switch	56 (81%)	207 (74%)	206 (56%)
Match	13 (19%)	73 (26%)	165 (44%)

$$\chi^2(2) = 32.42 \text{ } p<.001, \text{ Cramér's V=.21}$$

b) Full data set

	full pronoun	clitic pronoun	demonstrative pronoun
Switch	159 (92%)	506 (85%)	206 (30%)
Match	14 (8%)	93 (15%)	481 (70%)

$$\chi^2(2) = 479.52 \ p<.001, \ \text{Cramér's V}=.57$$

Comparing switch-match ratios horizontally, it can be seen that the prediction is not borne out. Rather, clitic personal pronouns show a weaker tendency for switching than their full form counterparts. Demonstratives, in turn, are even less frequent switchers. This pattern can be seen in both conditions. Cramér's V of more than 0,5 shows that the effect is particularly large in the full data set.

The reasons for the huge difference between the full and the restricted data is particularly due to the fact that the extended set includes common-to-common matches such as *de trein* [C] - *die* [C] 'the train - that'. This combination is extremely frequent; it makes up nearly a third of all the pronouns in the subcorpus. If the demonstratives are removed from the calculation, there remains only a weak correlation between the form of the pronoun and the likelihood of switching ($\chi^2(1) = 6.21 \ p<.1$, Cramér's V=.09).

Possible reasons for the differences in switching behaviour of the three pronoun types are discussed in section 8.4 below.

Next, we wanted to find out whether the Dutch data shows sensitivity to the case of the pronoun. Earlier research had raised the expectation that, when there is a choice between semantic agreement (switch) and syntactic agreement (match), "the likelihood of semantic agreement in the nominative will be as high as or higher than the likelihood of semantic agreement in the oblique cases" (Corbett 1991: 238). The hypothesis was first tested for personal pronouns and demonstrative pronouns together.

(7) Nominative versus oblique: personal and demonstrative pronouns

a) Restricted data set

	Nom pron	Obl pron
Switch	282 (70%)	187 (59%)
Match	122 (30%)	129 (41%)

$$\chi^2(1) = 8.81 \ p<.01, \ \text{Cramér's V}=.11$$

b) Full data set

	Nom pron	Obl pron
Switch	497 (61%)	374 (58%)
Match	314 (39%)	274 (42%)

$$\chi^2(1) = 1.9 \text{ p<.5, Cramér's V=.04}$$

In both conditions, the corpus results confirm the hypothesis. Nominative pronouns show a higher propensity to switch than oblique pronouns. Statistical significance is weak, but established, and the effect is slightly more marked in the restricted data set.

The question was raised whether the pattern is indeed connected to nominative vs. oblique case marking or rather to subjecthood vs. objecthood, as claimed by Ihalainen (1985: 161, 1991: 115), Wagner (2003: passim) and Siemund (2008: 58). Dutch offers a window on this issue. If the effect is as strong in the demonstrative pronouns, which do not mark case morphologically, as in the personal pronouns, which do, then the reasons are likely to be syntactic. If, by contrast, the effect is prominently due to the case-marked personal pronouns, then morphology is expected to have an influence. Thus, the test was repeated for personal and demonstrative pronouns separately. Again, the results are given for both data sets.

(8) Nominative versus oblique or subject versus object: personal pronouns

a) Restricted data set

	Nom pron	Obl pron
Switch	166 (82%)	97 (64%)
Match	33 (18%)	53 (36%)

$$\chi^2(1) = 16.19 \text{ p<.001, Cramér's V=.22}$$

b) Full data set

	Nom pron	Obl pron
Switch	381 (90%)	284 (82%)
Match	43 (10%)	64 (18%)

$$\chi^2(1) = 10.89 \text{ p=.001, Cramér's V=.12}$$

Excluding the demonstrative pronouns lowers the number of cases considerably, but the effect is clearly more marked than in the combined set of personal and demonstrative pronouns. Note that the switch/match ratio shifts as the inclusion of common-to-masculine adds a substantial number of switches while matches can

only be increased by neuter pronouns (since there is no common gender personal pronoun).

Interestingly, if we turn around the set-up of the database query and only consider the demonstrative pronouns, the effect disappears.

(9) Nominative versus oblique or subject versus object: demonstrative pronouns

a) Restricted data set

	Subj pron	Obj pron
Switch	116 (54%)	90 (51%)
Match	89 (46%)	76 (49%)

$\chi^2(1) = 0.21$, p<1 (not significant)

b) Full data set

	Subj pron	Obj pron
Switch	116 (43%)	90 (40%)
Match	271 (57%)	210 (60%)

$\chi^2(1) = 0.0$, p<1 (not significant)

When the demonstratives are considered in isolation, there is no significant correlation between subjecthood or objecthood of the pronoun and its likelihood for semantic or syntactic agreement. This holds for both data sets. The effect measured in (7) above is therefore largely due to the personal pronouns. These differ from the demonstratives in their overt case morphology. As a result, for the Dutch pronoun data, the imbalance in the agreement behaviour of subject and object pronouns is explained better in terms of case marking than as a consequence of their syntactic function. This confirms the initial hypothesis: nominative pronouns have a stronger tendency to agree semantically than oblique pronouns.

8.3.2 Distance

The next issue to be addressed is that of distance. Distance between agreement controller and agreement target is one of the predictors of the likelihood of semantic agreement in general (Corbett 1979: 220). For the Dutch data, two distance measures were taken. First, distance was evaluated numerally by counting the words between antecedent and pronoun. Word distance varied between 0 (adjacency) and 328 (the greatest encountered distance).

For word distance, switches were expected to score a higher distance count than matches. The numbers given in Table (10) apply to the restricted data set (no distance counts were made for the records that were only part of the extended data

set). A variance analysis (ANOVA) yielded the following results. Across the data set, there is a significant correlation between switch/match ratio and distance in words $(F(1, 806)=12.95, p<.001, r=.13)$. Switches - on average - are further removed from their antecedents than matches. The size of the effect is small $(r=.13,$ interpretation according to Field 2005).

(10) Mean distance in words

Switch/Match (number of cases)	Switch (N=497)	Match (N=313)
Sample mean of distance in words	22.07	11.0
Standard deviation	34.746	16.652

This shows that semantically agreeing pronouns are, on average, twice as far from their antecedents as syntactically agreeing pronouns. The general tendency supports the findings from earlier research.

One of the weaknesses of this rough calculation is that the numbers might be skewed by inter-pronoun effects. After all, many conversations contain several pronouns in a single agreement chain. It is conceivable that once a speaker has made the switch to semantic agreement, this choice may influence any subsequent pronouns for the same antecedent. In order to see if distance effects hold true regardless of this factor, the data was separated into *unique* switches/matches and *sequence* switches/matches. Unique switches/matches consist of an antecedent and a single agreeing pronoun, while for sequence switches/matches, the antecedent is followed by several pronouns. The unique cases provide evidence for 'raw' distance effects. In both unique and sequence cases, semantically agreeing pronouns are expected to be further removed from their antecedent than syntactically agreeing pronouns.

ANOVA indeed shows a significant interaction effect between switch/match ratio and uniqueness of the pronoun $(F(1, 806)=7.15, p<.01, r=.09)$. Thus, the statistics require a separate analysis for unique and sequence cases.

A calculation of the average word distance shows, again, that mean distances are higher for switches than for matches. This holds for unique pronouns as well as for pronoun in agreement sequences.

(11) Mean distance in words for unique and sequence cases

	Unique		Sequence	
Switch/Match (number of cases)	Switch (N=157)	Match (N=140)	Switch (N=340)	Match (N=173)
Sample mean of distance in words	6.01	4.11	29.49	16.58
Standard deviation	8.415	7.111	39.487	19.803

ANOVAs reveal that the effect is smaller for unique cases (F(1, 295)=4.39, p<.05, r=.12) than for sequence cases (F(1, 511)=16.38, p<.001, r=.18), but both are significant. Thus, the correlation between distance and likelihood of semantic agreement is stronger in pronoun sequences, but not entirely attributable to inter-pronoun influences such as persistence of an earlier choice.[1] Again, the corpus data supports previous hypotheses and findings.

As counting words does not take into account any syntactic or conversational structure, another independent distance measure was employed. Cases were sorted into four categories: *in clause*, *in sentence*, *in turn*, and *beyond*. Sentences were defined by the corpus transcription, just as were turns (including linearizing any parallel utterances by two or more speakers). All pronouns that were divided from their antecedents by a turn boundary were marked as "beyond".

Table (12) gives the numbers of switches and matches and their relative distribution for each distance measure. From what we know about the influence of distance, the switch/match ratio was expected to increase monotonically in favour of the switches, as we progress from *sentence* to *turn* and *beyond*. Unfortunately, "in clause" proved to be an unhelpful category. Due to their syntactic properties, pronouns cannot occur in the same clause as their antecedents, with the exception of possessive pronouns. Due to the scarcity of possessive pronouns in the corpus, the database only contains a single instance of a within-clause possessive, so the category was dropped.

For the other three measures, the result is as expected.

(12) The influence of syntactic distance

	In sentence	In turn	Beyond
Switch	115 (51%)	81 (62%)	300 (67%)
Match	112 (49%)	50 (38%)	151 (33%)

$$\chi^2(2) = 16.03, p<.001, \text{Cramér's V}=.14$$

Reading from left to right, the switch/match ratio indeed increases monotonically from sentence to turn and beyond. The effect is weak, but clearly significant. This implies that, again, greater distance favours semantic agreement. Note that the syntactic distance measure is independent of the word count measure: word count of zero (adjacency) can coincide with a turn boundary (and thus with the syntactic measure "beyond"), while sentences and turns can contain a large number of words, the extreme case being 40 words between antecedent and pronoun within a single sentence. Yet, both measures give the same result.

---

[1] Note that, strictly speaking, the first pronoun in a sequence should be counted as a unique choice because it is - by definition - not influenced by earlier choices. Such subtleties could not be accommodated for in the analysis.

Both word distance and syntactic distance were calculated as an average across the whole corpus. Yet, distance effects can also be witnessed on the level of the sentence. For every pronoun that was part of a larger agreement chain with other pronouns, the patterns of switching and matching pronouns were recorded. In this pool, there are only 53 cases where a semantically agreeing pronoun is followed by a syntactically agreeing one, against 460 cases where there are only switches, only matches or switches following matches. Thus, 'switching back' from semantic to syntactic agreement is the marked option. In many of the cases, the patterns reflect the pronoun choices of different speakers. For an individual speaker, to switch back is unusual.

8.3.3 The noun

Next in line, two factors related to the noun were investigated for their influence on pronominalization strategies. The first has to do with semantics.

The tendency to use a pronoun that is determined by semantics rather than syntax has been attributed to the fact that nouns can be felt as having the 'wrong gender', i.e. a gender that does not fit their meaning. It has been argued that pronouns are sensitive to their referent's degree of individuation. For highly individuated referents, common or masculine pronouns are used (feminine pronouns being restricted to female persons and - occasionally - animals) while low individuation favours neuter pronouns. If degree of individuation is viewed as a gradual phenomenon that can be expressed in a hierarchy, then the extreme ends of the hierarchy are the most conflictive. The more individuated a referent is, the more this clashes with neuter gender, while extremely low individuation conflicts maximally with common gender. The expectation is therefore that pronouns referring to entities at the extremes of the scale switch more readily than those towards the middle of the hierarchy. In order to see if this hypothesis is correct, switches and matches were sorted according to semantic class and the distribution was calculated separately for each class. Again, the test was run on both data sets. In both conditions, collectives, events/places, uncertain and ambiguous cases were omitted.

(13)

a) Restricted data set

	Person	Animal	Object	Specific mass	Unspecific mass
Switch	129 (94%)	13 (81%)	241 (52%)	22 (25%)	75 (88%)
Match	8 (6%)	3 (19%)	222 (48%)	65 (75%)	10 (12%)

$\chi^2(4) = 154.43$, p<.001, Cramér's V=.44

Table (13) shows that the prediction is indeed fulfilled. The likelihood of switching is much higher at the left and the right edge of the hierarchy and decreases in the middle (the standard residual is 5.0 and -6.2 for humans, -4.3 and 5.3 for specific

167

masses, and 3.2 and -4.0 for unspecific masses, which shows the strong contribution of these classes to the correlation). The specific mass nouns show the lowest propensity to switch. This can be attributed to their double status: they combine properties of objects and of substances and therefore do not give occasion to the feeling that a syntactically agreeing pronoun is semantically 'wrong'. Thus, their remarkably low percentage of switches can be accounted for.

The pattern is replicated in the full data set.

(14)

b) Full data set

	Person	Animal	Object	Specific mass	Unspecific mass
Switch	129 (94%)	50 (78%)	582 (52%)	43 (29%)	75 (69%)
Match	8 (6%)	14 (21%)	543 (48%)	107 (71%)	34 (31%)

$$\chi^2(4) = 154.15, p<.001, \text{Cramér's V}=.31$$

Again, the effect is strongly significant.

Last but not least, we were interested in the influence of the morphological structure of the antecedent noun. Many neuter gender nouns with non-neuter pronouns are diminutives, and it could be suspected that the gender of the root noun may influence the choice of the pronoun. To see if the data provides evidence for such an effect, all diminutive antecedents were tagged for their root gender, and the switch/match ratio was calculated for common and neuter roots. If the common gender of the root was of influence, the percentage of switches should be much higher in this group than in the group with neuter gender roots.

(15) Root gender of diminutives

	common gender root	neuter gender root
Switch	183 (78%)	24 (63%)
Match	50 (21%)	14 (37%)

$$\chi^2(1) = 4.29, p<.1, \text{Cramér's V}=.13$$

Table (15) does indeed show such a tendency, although the effect is weak. Whether it is indeed due to the gender of the root cannot be said with certainty.

8.3.4 Age of speaker

Besides linguistic factors, it has been suggested that the age of the speaker may influence his/her pronoun choice. This is because semantically agreeing pronouns are largely considered incorrect, especially for inanimate referents, and more

conservative speakers, which are often among the older generation, are probably more sensitive to such issues. Moreover, the impression is that the semantic system is gaining ground and thus asserting itself more strongly in the speech of the younger generations.

Since the metadata in the corpus permitted access to the speakers' ages, it was possible to sort them into age groups. Six such groups were distinguished. The numbers of switches and matches and the ratio between them is given in tables (16a) and b) below.

(16) Pronoun choice per age group

a) Restricted data set:

	≤ 20	21-30	31-40	41-50	51-60	>60
Switch	106 (72%)	293 (65%)	30 (54%)	22 (38%)	24 (52%)	21 (35%)
Match	41 (28%)	151 (35%)	22 (46%)	38 (62%)	23 (48%)	38 (65%)

$$\chi^2(5)=45.51, p>.001, \text{Cramér's V}=.24$$

b) Full data set:

	≤ 20	21-30	31-40	41-50	51-60	>60
Switch	174 (75%)	497 (58%)	44 (42%)	74 (48%)	47 (48%)	54 (41%)
Match	98 (25%)	362 (42%)	61 (58%)	80 (52%)	52 (52%)	78 (59%)

$$\chi^2(5)=34.79, p>.001, \text{Cramér's V}=.15$$

As expected, the freqency of switches decreases from more than 70% switches for teenage speakers to a mere 35% for speakers above 60 years of age. The increase of syntactic agreement is clearest from <20 to 40. Above that, the picture becomes more muddled. In the restricted data condition, there is a sharp increase of switches in the 50-60 group. In the full data condition, speakers between 40 and 60 display the same agreement preferences. Unfortunately, the subcorpus provides less evidence for the older generations. For the younger speakers, where plenty of cases are available, the results are most reliable. The general trend shows that younger people use semantic agreement more often than their parents, a sign that the new semantic system of gender agreement is indeed spreading in the spoken language.

8.3.5 Frequency

To conclude, some numbers should be given as to the general frequency of switched pronouns. Again, we can use two measures. For all nouns that trigger variation in the gender of their anaphors (the restricted condition), the following figures can be stated. The subcorpus of 500,000 words contains 810 relevant pronouns. Of these, 476 are switches, i.e. semantically agreeing pronouns. This means that 58% of the

pronouns do not have the 'correct' syntactic gender. If common-to-masculine switches are included, as well as nouns that do not trigger inconsistent agreements because their gender does not conflict with their semantics (the full condition), there are 1630 token pronouns, of which 899 agree semantically. Under this calculation, the switch/match ratio is 55% switches - still more than half of the cases.

Note that common gender person nouns and their pronouns were not counted. Their frequency is too high and their variability too low to make investigation worthwhile. Thus, the total number of pronouns in the corpus is higher than the present figures suggest.

8.4 Discussion: influences on variation and choice

Summing up the results of the corpus study, the following factors seem to be of influence on the choice for semantic or syntactic gender agreement in spoken Dutch.

The first factor is the pronoun itself. Personal pronouns are more likely to agree semantically than relative pronouns. This result is in line with the Agreement Hierarchy (Corbett 1979, 1991) which predicts that items lower on the scale should show a greater tendency to agree semantically.[2] Relative pronouns rank higher on the scale than personal pronouns. Possessive pronouns, which do not have a place of their own on the Agreement Hierarchy, show an even stronger tendency towards semantic agreement in the Dutch data, although the number of cases is too low to provide conclusive evidence.

Amongst the anaphoric pronouns, there are differences in agreement behaviour between full form (free word) pronouns, clitic pronouns and anaphoric demonstratives. Free pronouns have the strongest tendency towards semantic agreement, followed by clitic pronouns. For the demonstratives, the likelihood to agree semantically is only between 30% and 56%. This is not surprising, taken into account the fact the paradigm of the demonstrative pronouns has the same gender values as that of the attributive elements: the definite article and the adjective. Thus, there is no paradigmatic mismatch between the agreement controller and its demonstrative target. Unproblematic targets are expected to take part in the general changes and developments in pronominalization, but they are likely to do so more reluctantly. This is relevant for the issue of the spread of semantic agreement, addressed below.

The reason that full pronouns show such a strong tendency to switch probably lies in the fact that full form pronouns are used relatively more often for persons than for

---

[2] The qualification "lower" applies to the scale given in Corbett (2006) where the hierarchy is presented as attributive > predicate > relative pronoun > personal pronoun. Occasionally, e.g. in Corbett (1991) and Corbett (2000), the hierarchy is given with "<" signs, in which case the personal pronouns would occupy the highest position.

inanimate objects, and pronouns with person nouns as antecedents have a switching ratio of 94% (see Tables (13) and (14) above). This circumstance may contribute to the high percentage of switches among the full form pronouns.[3] The original hypothesis that free word pronouns are associated with more conservative speech and thus exhibit more syntactic agreement is not supported by the data.

Next, the case of the pronoun was shown to correlate with its agreement behaviour. Nominative pronouns are more likely to take semantic rather than syntactic agreement compared to oblique pronouns. This result was not unexpected: earlier research has shown that, when there is a choice between agreement options, nominative forms are more likely to agree semantically.

The best-known influential factor in agreement choices is certainly the distance between the agreement controller and the agreement target. Especially when the targets are pronouns, which can be widely separated from their antecedents, such that agreement chains stretch across whole sections of dialogue, the issue is not unproblematic. For this reason, two independent distance measures were tested: linear distance in words and structural distance with the units "clause", "sentence", "turn" and "beyond". Both measures gave the same result: higher distance correlates with a higher percentage of semantic agreement. This again was as predicted by earlier findings. For the Dutch data situation, influences among several agreement targets in a chain were suspected, which made it desirable to check the distance effect separately for unique matches/switches (one pronoun per token antecedent noun) and sequence matches/switches (more than one pronoun per token antecedent noun). For both cases, a robust distance effect emerged.

Moreover, the corpus level distance effects were corroborated on the level of the sentence. In the majority of cases, semantic agreement followed syntactic agreement in an agreement chain. The reverse often reflected different construals or different strategies by different speakers.

The antecedent noun was expected to exert influence on agreement choices in two ways. First, its place on the Individuation Hierarchy determines the degree to which a noun's semantics clashes with its grammatical gender. If switching is a strategy employed by speakers to avoid such clashes, semantic agreement should be more likely at the extreme ends of the hierarchy and less frequent towards the middle. The data show that this is indeed the case. The figures thus support an analysis in terms of conflict avoidance and semanticization of pronominalization. After all, rejecting a

---

[3] One could argue that the dependency works the other way around: the high percentage of person switches is due to the many full pronouns which are likely to switch. However, the full pronouns form a minority among the switched pronouns with person reference: the bulk of these pronouns are clitic forms and demonstratives. Thus, if there is an influence between the two factors, the strong preference for semantic agreement with person nouns is likely to influence the statistics for the full form pronouns rather than the other way around.

pronoun that is perfectly appropriate in syntactic terms is expected to happen due to a semantic reason.

Second, it appeared necessary to check if the morphology of the noun could also exert influence on the behaviour of its pronouns. In particular, the issue was raised if pronouns with diminutive antecedents had the 'wrong' gender because they actually agreed with the gender of the noun root. The data shows that, indeed, diminutives with common gender roots were more likely to be pronominalized by common gender pronouns than diminutives with neuter gender roots.

Last, but not least, is was investigated if the age of the speaker correlated with his/her agreement choices. A general tendency was found for younger speakers to choose semantic agreement more often, while older speakers preferred syntactic agreement in a higher number of cases.

8.5 Spread

The figures on speaker age and agreement preferences suggest that the likelihood of semantics-based pronoun choice is increasing as the linguistic baton is passed on to younger speakers. The impression is supported by evidence from the relative pronouns.

In this book, it has been assumed that the rise of semantic agreement is connected to the paradigmatic mismatches that have uprooted the traditional Dutch gender system (Chapters 3 and 6). While there probably has been some degree of semantically motivated pronoun choice throughout linguistic history (after all, it can also be witnessed in gender systems with matching agreement targets such as in German), semantic agreement is now so widespread that it is often the rule rather than the exception (although it is still far from being accepted by the official grammar rules). Such massive switching does not occur in systems like German, where definite articles and pronouns distinguish the same gender value and where thus a syntactically agreeing pronoun is always at hand. For Dutch, the pronouns fail to line up with the other agreement targets in terms of their feature values, and it is likely that the new distribution of the pronouns is a consequence of this fact.

If this is the right approach, then the question arises why the relative pronouns participate in the switching. After all, their paradigms are reduced in the same way as that of the definite articles and the adjectives. All three categories distinguish common and neuter gender and have lost all traces of a masculine-feminine distinction (at least, in the northern standard variety of Dutch that is considered here). Simply speaking, relative pronouns have no reason to switch. In fact, it has been shown that they switch to a lesser degree than personal pronouns and anaphoric demonstratives do.

In my opinion, semantic agreement on relative pronouns is a sign of spread of the new semantic gender agreement rules from the personal pronouns to other targets. Such a spread, and the direction from personal pronoun to relative pronoun and

potentially onwards, is a common phenomenon in the diachrony of gender systems. Corbett (1983) desribes the spread of semantic agreement in Slavic, and Corbett (1991: 248) gives other examples. In the cases discussed there, the spread of semantics-based innovations is connected to the Agreement Hierarchy. They originate in the personal pronouns and spread towards targets lower on the hierarchy, such as the relative pronoun. In this light, the Dutch patterns are not unusual. Rather, the observations fit in with existing research on changes in agreement systems.

8.6 Within-NP agreement

This chapter should be concluded with a brief note on agreement within the NP. A separate search was conducted on a part of the Corpus Gesproken Nederlands, which comprised about a million words of spontaneous speech by speakers of standard Dutch. The search was done automatically, so the results depend on the correct tagging of the corpus data. The search items were the combination of a neuter gender definite article (*het* or *'t*) with a common gender noun and a common gender article (*de*) with a neuter gender noun. In order to facilitate searching, only adjacent determiner-noun pairs were considered. The search yielded 20 common gender nouns with neuter gender articles and 53 neuter gender nouns with common gender articles. These figures in no way correspond to the frequency of semantic agreement in the pronouns, despite the fact that for the articles, the subcorpus was twice as large. Thus, as far as frequency is concerned, diverging genders in the definite articles are a rare phenomenon. Moreover, it is impossible to discern any semantic patterns in the cases. Thus, we find common gender articles for count nouns such as in *de fototoestel* 'the camera' or *de touw* 'the rope', but also for mass nouns such as *de water* 'the water' or *de graan* 'the grain'. Conversely, there are erroneous instances of neuter gender articles for mass nouns, such as *'t sneeuw* 'the snow' or *'t spraak* 'the speech', but also for count nouns, as in *'t kattenbak* 'the cat's litter box' or *'t maand* 'the month', and even for human referents (*'t persoon* 'the person'). Thus, gender 'switches' in the attributive domain display all the characteristics of slips of the tongue and are quite dissimilar to the frequent and systematic switching of the personal pronouns.

8.7 Conclusion

This concludes the investigation of Dutch pronoun usage. It is argued that speakers employ two systems of pronoun usage, one syntax-based, the other semantics-based. Preference for one system or another is influenced by a number of factors. A statistical analysis of the data shows that the following factors correlate significantly with the choice between switch and match: the sort of pronoun (relative versus personal pronoun, full form versus clitic versus demonstrative), the case of the pronoun (nominative versus oblique), the semantics and the morphology of the noun, and the distance between noun and pronoun. Moreover, there seems to be a constraint on 'switching back' once a semantically agreeing pronoun has been chosen. Additionally, the age of the speaker was shown to be relevant.

With the help of the Individuation Hierarchy and the Agreement Hierarchy much of the variation can be accounted for, either in terms of conceptual construal, or as a competition between two alternative agreement systems.

In Chapters 3 and 6, it has been argued that the semantic system, the younger of the two, is an innovation following the merger of the masculine and the feminine in the attributive domain. When the use of the two associated pronouns ceased to be governed by syntax, a new division of labour arose amongst the pronoun genders. In the last part of this book, we will move on to other languages which show similar problems or similar behaviour. The first expedition (Chapter 9) will be a cross-Germanic journey, the second takes a wider look at other language families across the world (Chapter 10).

Part III
Beyond Dutch

Chapter 9

Diachrony
and Parallels in Germanic

The Dutch situation as described in the preceding chapters is neither diachronically nor genetically isolated or typologically exceptional. Both the loss of the gender markers and the resemanticization of the pronouns have parallels within the language family. In this chapter, the developments will be placed in a cross-Germanic context, both diachronally and synchronically.

9.1 Diachrony: patterns of gender loss

We saw in Chapter 3 that the linguistic situation in the contemporary Dutch gender system has its roots in language history. The reduction of the original three genders to two in all agreement targets but the pronouns gave rise to a paradigmatic mismatch. This reduction was caused by erosion which levelled the distinctions between genders. In Indo-European languages and in inflecting languages in general, erosion of gender markers is part of a wider development known as *deflection*, the reduction and loss of inflectional markers, which are often portmanteau morphemes, with the result that gender, number and case marking are affected in parallel.

An interesting observation made in the Indogermanist and typological literature is that gender reduction or loss proceeds in cross-linguistically predictable ways. As outlined in Chapter 3 above, the nouns themselves are the first elements to lose their overt gender distinctions if changes such as stress shift cause phonological merger of the forms. Weak formal oppositions are then gradually lost.[1] This deflection or erosion process may spread and affect other elements. Next in line are adjectives, which, according to Priestly's comparative Indo-European study (1983), tend to lose their inflections after the nouns, then followed by other attributive elements such as definite and indefinite articles. Distinctions are retained longest on agreement targets outside the noun phrase, particularly on personal pronouns (Corbett 1991: 143). The same observation is made by Marchese 1988 in a comparative survey of Kru languages, a genus of the Niger-Congo language family. While the available

---

[1] As many other sources, Priestly stresses that this process of phonological merger only results in gender loss if a weak formal distinction goes hand in hand with an unclear or inconsistent semantic opposition. On the other hand, syncretism itself can cause semantic oppositions to become opaque.

evidence is scattered across language families (see Corbett 1991: 143 and Corbett 2006: 274 for references), there is some support for a cross-linguistic hierarchy of the relative order in which agreement markers are reduced. The scale can be formulated as under (1).

(1)   noun > attributive > personal pronoun

This hierarchy is reminiscent of the Agreement Hierarchy, and indeed, we might expect that the relative pronoun would take its place between the attributive elements and the personal pronouns. Unfortunately, relative pronouns are cross-linguistically rare, so the typological evidence is lacking. With regard to the predicate, a study of various Cross River languages of the Niger-Congo branch of the Niger-Kordofanian phylum lends support to the assumption that predicates lose their gender agreement after numerals and adjectives (Demuth, Faraclas and Marchese 1986). This is in line with the Agreement Hierarchy. The most relevant generalization for the purposes of the present study is cross-linguistically robust: attributive elements generally lose gender distinctions at an early stage, while pronouns retain them longest amongst the agreement targets.

9.2 Pronouns in contemporary Germanic

The diachronic tendencies outlined above cause expectations about the synchronic state of gender systems. In synchronic terms, the hierarchy predicts a smaller number of gender values for elements towards the left edge of the hierarchy and a higher number for elements towards the right. Looking at the contemporary Germanic standard languages and comparing the gender values marked on the attributive elements with those marked on the pronouns, we see the expectation confirmed: seven of the twelve languages have a higher number of pronominal genders (Table 2). The typical sitation is that the attributive targets merge masculine and feminine to common gender.[2] Yet, all languages have preserved the original triad of masculine, feminine and neuter in the pronouns.[3]

---

[2] Contrary to Germanic, the Romance languages often show a collapse of masculine and neuter. Thus, modern Italian, French and Spanish only have masculine and feminine nouns. In all cases, the neuter pronouns have survived in a rudimentary fashion as neutral agreement markers (in the sense of Corbett 1991: 159) or expletives. A real exception among the Germanic varieties is the Frisian dialect of Fering-Öömrang which shows systematic syncretism between the feminine and the neuter. Yet, even in this particular case, there are still three pronominal genders, masculine, feminine and neuter (Hoekstra 1996).
[3] In some cases, the three pronominal genders have been joined by a fourth, often referred to as common or uter gender. Example forms are the Swedish pronoun *den* [C] and the Dutch demonstrative anaphors *deze* [C] or *die* [C] discussed above.

(2) Germanic attributive and pronominal genders

Language	Attributive genders			Pronominal genders		
a) German	masculine	feminine	neuter	masculine	feminine	neuter
b) Yiddish						
c) Icelandic						
d) Faroese						
e) Norwegian[4]						
f) Norwegian	common		neuter			
g) Swedish						
h) Danish						
i) Dutch						
j) Frisian						
k) English	–					
l) Afrikaans						

It is surprising that this state of affairs has attracted so little attention in linguistic literature. We have seen some of the theoretical problems in Chapters 1 and 2. First, many researchers do not acknowledge pronouns as agreement targets and thus would be hesitant to compare and contrast adnominal and pronominal gender. Else, if the pronominal genders are set apart for theoretical reasons, a mismatch between pronominal and adnominal gender values may not be regarded as problematic.[5]

---

[4] Norwegian has several standard varieties that display different degrees of syncretism between masculine and feminine. A standard reference grammar (Strandskogen and Strandskogen 1995) implicitly illustrates the difficult situation by maintaining that Norwegian has three genders, masculine, feminine and neuter, but showing that feminine forms can often be replaced by a masculine (or rather: common gender) form: the masculine indefinite article *en* can be used instead of the feminine *ei* (46), the masculine definite determiner *-en* (suffixed to the noun) can replace the feminine *-a* (62) and the demonstrative, used as the free definite article in front of adjectives, has the same form for both genders (46). The difference between the systems does not correspond to the two standards Bokmål and Nynorsk. While Nynorsk has a classical Germanic three-gender system in the articles and suffixes, different versions of Bokmål take different stances in the gender question. Originally, Bokmål propagated a two-gender system, but later the grammatical feminine was partly reintroduced (Duke, to appear). As is often the case, the issue is political as well as linguistic. As far as the dialects are concerned, the Bergen dialect seems to be predominantly two-gender, while many Eastern and Western Norwegian dialects still have three genders (Hans-Olav Enger, personal communication).

[5] Yet, as mentioned in Chapter 2, one of the reasons to consider pronouns as agreement targets much like attributive elements is the fact that they often distinguish the same gender values as the other agreeing elements in the language.

Second, the systems between f) and l) above violate the general expectation that agreement controller and agreement target should share not only the same feature (this may indeed be regarded as a condition for agreement, see Corbett 2006: 7), but also the same value range of this feature. When there are differences in value ranges, the theoretical as well as descriptive problems are so difficult that it may be tempting to split off the pronominal genders and treat them under a different heading. On the other hand, the high concentration of overdifferentiated pronominal genders in the Germanic standard languages makes it worthwhile to investigate the question of how related languages solve a shared problem.

In the following sections we will have a look at those languages in Table 2 that display a mismatch between the attributive and the pronominal genders. We will start with the two languages where gender is exclusively marked on the pronouns: English and Afrikaans. Then we will move on to the languages that are closer to Dutch in having two attributive and three or four pronominal genders. For each language, a brief review of the history of the gender system is given, followed by a short outline of the synchronic data situation. The main point of the following sections is to highlight the way in which each language has solved the mismatch problem and reorganized the distribution of personal pronouns.

9.3 Purely pronominal gender: English and Afrikaans

9.3.1 English

Old English possessed a gender system comparable to that of contemporary German. Grammatical masculine, feminine and neuter were marked on articles, adjectives, numerals as well as relative and personal pronouns. Moreover, there were a few gender-specific inflectional suffixes on the nouns themselves. According to Lass (1992), the system shows signs of decay since as early as the 10th century, judging from gender vacillation. Around 1200, the system is "in considerable disrepair" (107). Progressive erosion and syncretism of inflectional morphology stripped most agreement targets of their gender. The personal pronouns remained unaffected by this development. As the last gender-marking elements (the opposition between *what* and *who* not usually being regarded as a gender distinction, although Quirk et al. 1985: 341 suggest such a view), they even show an increased formal distinctiveness between the genders by the introduction of the feminine pronoun *she* instead of the older form *heo* that was near-syncretic with the masculine pronoun (Curzan 2003: 45).

Gender assignment was of the mixed type consisting of formal and semantic rules. Semantic rules basically applied to nouns referring to persons and were sex-based (with well-known exceptions such as *wif* [N] 'woman'). Formal rules were connected to inflectional classes. For example, *u*-stem nouns such as *duru* [F] 'door' were non-neuter.

Already in Old English, pronouns could defy this system. Thus, we find cases of syntactic agreement side by side with semantic agreement, as in (3).

(3) Old English

þæt	þu	þone	**wisdom**	þe	þe	God	sealde,
that	you	that	wisdom	which	to.you	God	gave

þær	þær	þu	**hiene**	befæstan	mæge,	befæste.
there	where	you	3.M	implant	may	implant

Geþenc	hwelc	witu	us	þa	becomon
think	what	punishments	to.us	then	came

for	þisse	worulde,	þa	we	**hit**	nohwæþer
for	this	world(F)	when	we	3.N	neither

ne	selfe	ne	lufodon,	ne	eac	oþrum
not	ourselves	not	loved	not	as	other

monnum	ne	lefdon
men	not	allowed

'… that wisdom which God gave to you, where you may implant it, there implant it. Think what punishments would come to us for this world if we did not love it nor allowed others to do so …' (*Preface* to the *Cura Pastoralis*, quoted from Dekeyser 1980: 101)

Whether Old English speakers preferred syntactic or semantic agreement is contested in the literature. Older studies, e.g. Moore (1921), claim that the use of the pronouns in Old English "was almost the same as our own use of them" (89). However, there is a danger that these studies overrate the pronouns for person reference, which constitute the bulk of the pronouns in the corpora. For pronouns referring to inanimate nouns, Curzan (2003), an extensive diachronic study, reports "robust health" (91) of the traditional system of grammatical gender agreement until the Early Middle English period.

Between Middle English and Modern English, however, the original system was abandoned in favour of the purely semantics-based pronominal gender system of today (see Siemund 2008: 9-12 for an overview of the literature on this development). Clearly, then, while history did not reduce the form inventory of the personal pronouns, the changes did have consequences for pronoun usage and the distribution of the pronominal genders. Whereas in Old English, pronouns generally agreed syntactically with their antecedent - at least in inanimate reference - in Modern Standard English they are distributed according to semantic considerations.

The semantic organization of the contemporary Standard English gender system is comparatively simple. Generally speaking, the feminine pronoun is used for female persons and occasionally for female animals, while the masculine pronoun appears in reference to male persons and a few male animals. All other referents are pronominalized with neuter gender. Such a system is often referred to as a "natural gender system" (Corbett 1991: 9). Exceptions are the often cited feminines for countries and ships, as well as occasionally for cars and machines. While this phenomenon cannot be discussed in much detail in this book, it is interesting in its own right.[6] In particular, the question can be raised if this sort of agreement is not actually a remnant manifestation of syntactic agreement with grammatical feminines, in much the same way as English has syntactic agreement with pluralia tantum such as *trousers* or *scissors*. These nouns are formally plural but semantically singular, yet they take plural agreements. The connection between ships and cars and femininity is not easy to motivate otherwise. Yet, most studies explain the choice of the feminine versus the neuter on semantic grounds. Indeed, uses of masculine or feminine pronouns for animals or inanimate entities are often discussed under the name of personification, either in the narrow sense (of humanizing animals, e.g. in fiction, see McKay and Konishi 1980) or in a wider sense (when inanimate entities are equipped with properties usually associated with personhood, without fully equating them with human beings). This approach is supported by observations such as that countries can only be pronominalized by a feminine pronoun when they are considered as political rather than geographical entities (Huddleston and Pullum 2002: 488). Note the compatibility with the *degree of individuation* approach outlined in Chapter 6.

Less generally known are dialectal varieties that show a different picture. Siemund (2002a,b and 2008) discusses the traditional dialects in the southwest of England (West Somerset English) as well as in Newfoundland (Paddock 1991, Wagner 2003) and Tasmania (Pawley 2002 and 2004). In all varieties, masculine forms are used for inanimate entities. Thus, we find masculine pronouns for the antecedents *hand, tree, letter* and *hat* (Siemund 2008: 43-44, dialectal spelling normalized). Neuter pronouns, in turn, are used for nouns such as *water, beer, coal* and *beef* (Siemund 2008: 44-46), unless a "specific piece of material is picked out" (46).

Though such usage may look like the remnants of a syntactic system, Siemund (following Elworthy 1886 [1965]) shows the clear semanticity of the pronominal gender system in the mentioned varieties. Besides distinguishing male and female persons, as in Standard English, the pronouns reflect the opposition between bounded or countable and unbounded or uncountable referents, just as in Dutch. Siemund (2008: 140 f) links these patterns to the Individuation Hierarchy (although he includes proper names as a separate category and sorts abstract referents between tangible objects and masses). Schematically, this gives the following patterns for the standard English and the Southwest English pronominal genders.

---

[6] Relevant data and discussion is, for example, provided by Morsbach (1926: 30), Svartengren (1927), Marcoux (1973) and Mathiot and Marjorie (1979).

(4) Individuation Hierarchy and English pronominal gender

	Proper names	Persons	Animals	Inanimate tangible objects	Abstracts	Mass nouns
Standard English		*she*		*it*		
		he				
Southwest English		*she*		*he*		*it*
		he				

We see that the two systems make different cut-off points on the same scale. For both variants, there is a separation between animates that are distinguished according to their natural gender. To the right of this boundary, Standard English has only one remaining class: the neuter pronoun is used for anything (ships and cars aside) from the rest of the animates down to the mass nouns. The Southwest English pronominal system makes a second cut-off point, separating inanimate tangible objects from abstracts and mass nouns. For the former, the masculine is used, while the latter is the domain of the neuter.

Summing up, the English system is strictly semantic and alignable to the Individuation Hierarchy.

9.3.2 Afrikaans

The situation in Afrikaans is much less clear and less well known. Afrikaans developed out of 17th century Dutch, a language that was on its way from a three-gender system to a two-gender system. In early Afrikaans there is evidence for two genders: the common gender with the definite article *de* and the demonstrative *die*, as well as the neuter gender with the definite article *het* and the demonstratives *dit* and *dat*. The complex gender assignment system of Dutch that assigns nouns to genders by a variety of rules did not survive at the Cape, one of the pressures probably being the complex language contact situation. By the end of the 18th century, the common gender had ousted the neuter and the former common gender demonstrative *die* had become the only definite article in the language (Scholtz 1963: 125, Ponelis 1993: 172, Deumert 2004: 106). Thus the language lost its nominal gender distinction.

As in English, however, the gender morphology was fully preserved in the personal pronouns. Modern Afrikaans has three personal pronouns, *hy* 'he', *sy* 'she' and *dit* 'it' (a former demonstrative). The masculine and feminine have the oblique case forms *hom* and *haar*. The few studies that exist about the usage of these pronouns suggest the following development.

Afrikaans inherited from Dutch the tendency to use the feminine pronoun almost exclusively for female animates (especially persons). Thus, the grammatical feminine had practically disappeared from the language, and the interesting changes

took place between the masculine and the neuter pronoun. In the 18th century, the neuter pronoun *dit* seems to have extended its domain into that of the masculine (Scholtz 1963: 132, 1966). Scholtz cites an unpublished dissertation that reports 208 neuter pronouns in anaphoric reference to inanimate entities, with only 14 masculine pronouns on the other side (1963: 133).

At the same time, Afrikaans started to combine former Dutch neuter nouns with masculine pronouns. Ponelis (1993) gives the following example:

(5)   Afrikaans
        Die bed is geskuif want   **hy** was in die pad
        DEF bed is moved because 3.M was in DEF way
        'The bed was moved because it was in the way'

For comparison, see the contemporary Dutch version (syntactic agreement):

(6)   Dutch
        Het bed   is verschoven want   **het** was in de   weg
        DEF.N bed(N) is moved   because 3.N was in DEF.C way(C)
        'The bed was moved because it was in the way'

In the 20th century, the tide turned in favour of the masculine. Ponelis (1979) notes that "*dit* continually retreats from the agressive advance of *hy*" (1979: 585, author's translation). This has the consequence that excessive use of neuter pronouns is increasingly perceived as "bookish" (Donaldson 1993: 127). According to a modern coursebook, the neuter pronoun *dit* and the masculine pronouns *hy* and *hom* (the oblique form) are "just as common" in inanimate reference (Donaldson 2000: 13).

In order to structure this confusing data situation, Ponelis (1979: 585 ff) identifies several different parameters that influence the decision between neuter or masculine. Next to the parameter style (formal versus colloquial), he mentions two conceptual parameters that make reference to the semantic content of the pronouns. These are [concrete] versus [abstract] and [countable] versus [uncountable], both familiar from the Dutch and the Southwest English cases. Combined into a hierarchy, these parameters form a continuum of inanimate entities between count nouns (Ponelis: "soortnaamwoorde"), concrete mass nouns ("konkrete massanaamwoorde"), abstract mass nouns ("abstrakte massanaamwoorde") and 'pure abstracts' ("suiwer abstrakta"). The pure abstracts are sentences and nominalizations which semantically speaking have got little in common with nominal reference.

This gives the following distribution of masculine and neuter pronouns (Table 7). Examples illustrate the slightly unusual semantic classification. The relative frequency for neuter and masculine increases in favour of the neuter, as we move from left to right in the table. The decrease of the masculine is slower in colloquial speech than in (formal) writing.

(7) Masculine and neuter pronouns in Afrikaans

hy			*dit*
←			→
count nouns	concrete mass nouns	abstract mass nouns	pure abstracts
rivier 'river'	*olie* 'oil'	*geduld* 'patience'	*hantering van vrag* 'handling of freight'
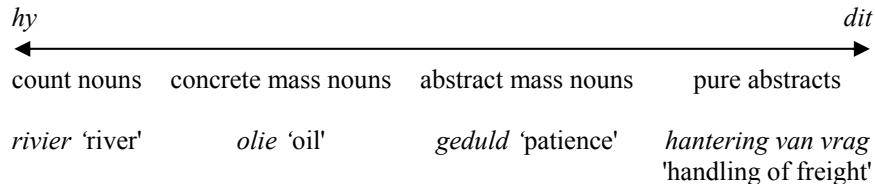
Besides, Ponelis stresses that affective load influences the choice of pronoun gender. "Dramatic, lively language use" (Ponelis 1979: 587) increases the chances that a masculine will be used for a mass noun, for which the neuter is otherwise the unmarked choice. By contrast, the masculine is the neutral option for count nouns. Ponelis interprets the situation as a steady advance of the masculine into the domain of the neuter. Today, the dominance of English at the Cape may tip the scales in the opposite direction, favouring neuter pronouns for inanimate referents. Whichever way the development proceeds, the heartland of the masculine lies in the left part of the hierarchy, that of the neuter in the right. This suggests that pronominal gender is, again, sensitive to referent individuation. Independent of Ponelis' work, this interpretation is also suggested in Siemund (2008: 185 f).

Generally, Afrikaans and English both show a new motivation for the use of anaphoric pronouns. Both systems are semantically organized, and can be linked to a conceptual hierarchy that separates countable, concrete elements on the one hand from abstract and/or mass entities on the other.

Next, the question arises what developments can be seen in those languages that have reduced their nominal genders without losing them entirely. This is the case in Dutch, as we have discussed in detail, but it also holds for all three mainland Scandinavian languages (Danish, Swedish, Norwegian) and for Frisian. The next section addresses the question of how Scandinavian and Frisian have coped with the mismatch situation. Particular attention will be paid to those genders that are exclusively marked on personal pronouns.

9.4 Partly pronominal genders: Scandinavian and Frisian

9.4.1 Scandinavian

The history of the Scandinavian genders is marked by strong regional and dialectal variation in the move from the traditional Germanic three-gender system to the two-gender system that now prevails on the Scandinavian mainland (for an overview of the historical paradigms and the present situation see Bandle et al. 2002 and Braunmüller 2000). Today, Scandinavia has varieties that have three genders throughout them (the German type), others that resemble the Dutch type, and a tiny minority that looks like English with only pronominal gender. The island languages Icelandic and Faroese, as well as some varieties of Norwegian and - more rarely - Swedish and Danish distinguish three nominal genders, while Standard Danish and

Standard Swedish as well as varieties of Standard Norwegian only have two nominal genders, common and neuter. In Denmark, there is Jutish, which is said to have lost the attributive gender marking altogether and only possesses pronominal gender (Bechert 1982: 27; Braunmüller 2000).[7] In all languages of the mainland, however, the pronouns distinguish four genders: masculine, feminine, common and neuter. The common and the neuter gender are represented by the two former demonstratives *den* and *det*. In the three standard languages, the masculine and the feminine gender are exclusively pronominal. The nominative singular forms are *han* for masculine (in all three languages) and *hon* (Swedish) respectively *hun* (Danish and Norwegian) for feminine gender.

As regards pronoun usage, all three standard languages display a split between animate and inanimate referents. The purely pronominal genders, masculine and feminine, are restricted to animate reference, specifically in use for persons and higher animals. Here, the syntactic gender of the antecedent noun is less relevant: neuter nouns referring to persons such as Swedish *barn* 'child', *biträde* '(shop) assistent' or compounds with *-råd* '-minister, secretary in civil service' are pronominalized not with a neuter pronoun, but according to the natural gender of the person. This also holds for the gender marking on predicative adjectives, as in the following example (Holmes and Hinchcliffe 2003).

(8) Swedish

a) **Det** unga **statsråde-t** var **säker** på sin sak.
DEF.N young cabinet_minister-DEF.N was sure.C of her case

Hon hade läst på.
3.F had read about

'The young cabinet minister was sure of her case. She had read up on it.'

b) **Affärsbiträde-t** blev **orolig**. **Han** hade stulit pengar.
shop_assistant-DEF.N became uneasy.C 3.M had stolen money
'The shop assistent grew uneasy. He had stolen money.'

Thus, for animate referents, pronoun choice is on semantic grounds. For inanimate referents, the choice of either *den* or *det*, the common respectively neuter anaphoric pronoun, is determined by the syntactic gender of the noun. See (9) for two examples from Danish (Allan et al. 2000:155).

(9) Danish

a) Hvor er **bog-en**? Jeg lagde **den** på bordet.
where is book-DEF.C I put 3.C on table(N)
'Where is the book? I put it on the table.'

---

[7] West Jutish seems to have reinstated the attributive gender with the help of the demonstrative pronouns (Ringgaard 1973, Bechert 1982: 27; Braunmüller 2000).

b) **Bord-et**   passer ikke til   stuen;   **det** har ikke den
   table-DEF.N fits   not   with room(C) 3.N has not   DEF.C

   rigtige   farve.
   right   colour(C)

   'The table doesn't match the room; it hasn't got the right colour.'

It is interesting to note that this also holds for those varieties of Norwegian where masculine and feminine articles still exist. Although this suggests an intact knowledge about masculine and feminine gender, inanimates are pronominalized not by the masculine *han* or the feminine *hun*, but by the common gender *den*. An example from Strandskogen and Strandskogen (1995: 105) is (10).

(10) Der   ligger det   en **bok**.   **Den** er min.
     there lies   there a   book(M). 3.C is mine
     'There is a book there. It's mine.'

This fact has consequences for our understanding of the causality between gender loss and redistribution of pronouns, as it suggests caution about the common assumption that loss necessarily precedes redistribution. The issue is discussed in section 9.6 below. An interesting and extensive study about the gradual replacement of *han* and *hun* by *den* in Swedish is Davidson (1990).

More conservative varieties that still possess syntactically agreeing masculine and feminine pronouns for inanimate entities, are in the minority in Scandinavian as spoken today.

Besides, some dialects show semantics-based pronoun usage for inanimate referents very similar to that described for the South English varieties. In 6.4 we saw West Jutish, which marks the opposition between countable and uncountable entities on pronouns and articles, with pairs such as *den træ* 'that.C tree(C)' and *det træ* 'that.N wood(N)' (see Ringgaard 1973, Wahrig-Burfeind 1989, Gachelin 1991 or Allan et al. 2000).

Returning to the general picture, the Scandinavian languages display a split in the pronominal genders. From the point of view of the noun, this split separates animate from inanimate entities: for the former, the gender corresponds to the sex of the referent, for the latter, the pronoun takes the lexical gender of the noun. The situation is schematized in (11).

(11) Nominal and pronominal gender in Scandinavian

Semantics		Lexical gender of the noun	Pronominal Gender
Inanimate		*common*	*common*
		neuter	*neuter*
Animate	male	*common*	*masculine*
		neuter	*masculine*
	female	*common*	*feminine*
		neuter	*feminine*

Turning around and regarding the situation from the point of view of the pronouns, the facts can be stated differently. The four pronominal genders fall into two groups, one that has a counterpart in the attributive paradigms (common and neuter gender) and one that is purely pronominal (masculine and feminine). Syntactic and semantic agreement follows exactly this split. Those pronouns that have an attributive counterpart agree syntactically, while those that are not marked attributively are distributed according to semantic reasons. Of course, this is partly circular: if attributive agreement is taken as indicative for the gender of the noun, then a pronoun must have an attributive counterpart in order to be recognized as syntactically agreeing. Yet, in another perspective, the match is non-trivial. On the one hand, common and neuter gender pronouns could be used according to semantic considerations (as we have seen for Dutch). Thus, semantic agreement does not require a featural mismatch. On the other hand, masculine or feminine pronouns could theoretically agree syntactically, say with the gender that the nouns had in earlier times when the distinction was still alive in the language. This definition of syntactic agreement does not rely on an overt feature match between attribute and pronoun. If two nouns with the same attributive gender take different pronouns, say, if *bog* 'book' in Danish is pronominalized with a feminine pronoun, while *bil* 'car' takes a masculine pronoun, and if there is no discernible semantic reason for this fact, then this could be analyzed as syntactic agreement. It is significant that such patterns do not occur.

The issue will be taken up in section 9.5, after a brief look at Frisian.

9.4.2 Frisian

Frisian is genealogically as well as geographically a close neighbour to Dutch and strongly influenced by it. It is therefore not surprising that the gender systems of both languages are very similar. In most varieties of Frisian, the adnominal elements show syncretism of masculine and feminine. Again, the pronouns have retained the original three-way distinction between masculine, feminine and neuter. In one case, the dialect of Fering-Öömrang, it is the feminine and the neuter that have fused, a very rare situation in Germanic (Hoekstra 1996). Yet, even here the pronouns still have three genders. This fact is noteworthy as the fusion seems to have affected the pronouns, too: the former neuter pronoun *hat* takes over the function of the feminine. However, a new pronoun *det* (which according to Hoekstra 2006 stems

from the article system, although the demonstratives also come to mind as a likely source) takes the place of the neuter pronoun and thus reinstates the complete feature paradigm with three gender values.

From the little material available, some relevant data can be reported from the varieties of Town Frisian and Helgoland-Frisian. The comparative study by Wahrig-Burfeind sketches the following patterns. In Town Frisian (*Stedsk*), the masculine and feminine pronouns *hij* and *sij* are used according to the natural gender of the referent, while "for nouns that denote substances and are not neuter, [...] the personal pronoun 'it, 't is used" (Wahrig-Burfeind 1989: 176, author's translation). Wahrig-Burfeind stresses that this usage is a direct consequence of the syncretism in the (ad)nominal genders. Moreover, the pronoun *dy* is mentioned, a common gender demonstrative similar to Dutch *die*, which is used for "non-neuter object nouns" (Wahrig-Burfeind 1989: 178 quoting Fokkema 1967). Here, again, we see the double semantic split into male/female animates and countable/uncountable inanimates, together with some degree of syntactic agreement for *dy* and '*(i)t* (which is not explicitly given in the source, probably because it is the expected situation). The same holds for Helgoland-Frisian where the feminine pronoun *dji* (and its oblique form *her*) is used for female persons, the masculine *hi/hem* for male persons and countable referents, while the neuter pronoun *deät* appears in reference to mass nouns (Wahrig-Burfeind 1989: 199).

Schema (12) shows that the situation in Frisian is indeed similar to Dutch. Semantic agreement is found for the two purely pronominal genders, the masculine and the feminine, and it can overrule syntactic agreement in reference to mass nouns. In Helgoland Frisian, there is also semantic agreement for inanimate countable referents, even if the noun is neuter.

(12)

	Person	Animal	Inanimate object	Mass
Town Frisian	*sij* *hij*	*dy*	*dy/'(i)t*	*'(i)t*
Helgoland Frisian	*dji* *hi*	*hi*	*hi*	*deät*

Again, the patterns correspond to the conceptual hierarchy that proved useful for Dutch, English and Afrikaans.

9.5 Synthesis: the semanticity of pronominal genders

The cross-Germanic facts paint an interesting picture of the patterns of gender reduction in relation to the distribution of pronominal forms. Germanic pronouns can agree syntactically, that is, with the lexical gender of the noun and independently of its semantics. This is the case with the common and the neuter pronouns in Scandinavian and optionally with the common and the neuter pronouns

189

in Dutch and Frisian. Elsewhere, pronouns show semantic agreement. In Modern Standard English and in Afrikaans, this is the only type of agreement found.

Scandinavian on the one hand and Dutch and Frisian on the other differ in the distribution of syntactic and semantic agreement. In Scandinavian, semantic agreement is largely restricted to the animate domain, particularly to person reference. Inanimate referents are pronominalized according to the lexical gender of the antecedent noun. In Dutch and Frisian, by contrast, there is an across-the-board choice between syntactic and semantic agreement in both animate and inanimate reference.

The most interesting observation from the cross-Germanic data is that syntactic agreement only appears with pronouns whose gender value is also formally marked on attributive elements. This fact can be captured in the following generalization.

(Hypothesis A) All pronouns that agree syntactically must have an attributive counterpart with the same gender value.

Unfortunately, in this formulation the statement is circular: it trivially emerges from the definition of syntactic agreement. If syntactic agreement means agreement according to the lexical gender of the noun, and if the lexical gender of the noun is read off from the attributive elements, then coherence between attributive and pronominal gender is a necessary condition for syntactic agreement. Yet, syntactic agreement can also be viewed as agreement independently of semantics. This view suggests a different variant of the hypothesis. It could be formulated as follows.

(Hypothesis B) All genders that have personal pronouns as their sole formal exponent must agree semantically, i.e. their distribution must be organized according to semantic principles.

This generalization is testable and falsifiable. It would in principle be possible for nouns to have a lexical gender that is not marked on attributive elements but which determines pronoun choice. This is the option envisaged by the Dutch prescriptive grammars. The Dutch-Belgian committee that produced the first *Woordenlijst van de Nederlandse Taal* presumed that the threepartite distinction in the pronominal domain reflected the three historical genders of the nominal domain and advised that pronouns should be used accordingly (Verhoeven 1990: 495). Thus, the Dutch noun *machine* 'machine' to this day is considered a feminine noun by the official spelling dictionary (the most recent edition being its online resource http://woordenlijst.org), while the noun *berg* 'mountain' is called a masculine. This is not due to any semantic rule. Rather, the assumption is based on the fact that these words belonged to the feminine respectively masculine gender in the past. Thus, the authorities propose a situation where the two genders - the masculine and the feminine - are solely expressed through pronominal exponents and yet are realized by means of syntactic agreement.

The usage facts show that such a scenario is not in line with the cross-Germanic reality. In all relevant languages, purely pronominal genders are distributed according to semantic considerations. Table (13) illustrates the situation. Shaded cells indicate the correlation between the absence of an attributive marker and the agreement preference.

(13) Syntactic and semantic agreement in Germanic

	Attributive genders	Pronominal genders	Syntactic agreement	Semantic agreememt
English		*neuter*		+
	-	*masculine*	-	+
		feminine		+
Afrikaans		*neuter*		+
	-	*masculine*	-	+
		feminine		+
Scandinavian	*common*	*common*	+	-
	neuter	*neuter*	+	-
	-	*masculine*	-	+
	-	*feminine*	-	+
Dutch/Frisian	*common*	*common*	+	+
	neuter	*neuter*	+	+
	-	*masculine*	-	+
	-	*feminine*	-	+

Lack of attributive marking for a particular gender always coincides with semantic rather than syntactic agreement.

The inability of personal pronouns to support a non-semantic gender has not received much attention in the literature. An early formulation of the relevant question can be found in Classen (1919) who discusses the change from grammatical to natural gender in Old to Middle English and notes:

> [T]he new natural gender must have been expressed by the available material, that is, without the help of any inflexions or suffixes. How, then, was it, in fact, expressed? It was expressed solely by means of the personal pronouns; for all other distinctions of gender had been lost. Is it not then a perfectly natural and obvious objection to this theory, that these self-same pronouns, *he*, *she* and *it*, would have sufficed to preserve the old grammatical gender? If, ex hypothesi, there were no other means of expressing gender than the pronouns *he*, *she* and *it*, would not these pronouns have served just as well to express grammatical gender as to express natural gender?" (Classen 1919: 98)

Most accounts formulate the loss of agreement morphology and the resemanticization of the gender system as a necessary consequence of the latter from the former. For example, Wahrig-Burfeind states that "syncretism of the nominal genders always triggers a change in the pronominal system" (1989: 175, author's translation). While this appears to be true, it is noteworthy how little self-evident the development is.

Naturally, the best way to investigate the typological validity of Hypothesis B is to look into languages that only have pronouns as gender markers. If they, like English and Afrikaans, are semantics-based across the board, we have good evidence for a typological rule. This will be the concern of Chapter 10. However, before embarking on this issue, we will have a brief look at the diachronic relation between the loss of gender markers and the resemanticization of the pronouns.

9.6 The time course of loss and gain

Are the new uses of pronominal gender an invention following attribute syncretism or are they merely the spread of a phenomenon that existed anyway? As this study could not incorporate historical evidence, primarily because of the lack of natural data, we will call on observations from modern language use. To this purpose, it is helpful to look at more conservative Germanic languages that still have a three-gender system in the attributive paradigms.

9.6.1 German

Historically, Old English already provides evidence that intact three-gender systems can show pronominal semantic agreement, at least for animate referents. The same can be observed in the relatively conservative Modern High German. This language has a three-gender system with agreements for masculine, feminine and neuter on the definite article, the adjective, the relative pronoun as well as the personal and demonstrative pronouns in anaphoric use. Moreover, the possessive pronoun marks gender, with syncretism in masculine and neuter, but a distinct form for the feminine.

(14) Gender agreement in German

	Definite article	Adjective suffix	Relative pronoun	Personal/ Demonstrative pronoun	Possessive pronoun
Masculine	*der*	*-er*	*der*	*er/der/dieser*	*sein/seine*
Feminine	*die*	*-e*	*die*	*sie/die/diese*	*ihr/ihre*
Neuter	*das*	*-es*	*das*	*es/das/dieses*	*sein/seine*

While the two attributive categories always agree syntactically with their noun, the relative, personal, demonstrative and possessive pronoun can show semantic agreement (relating to the possessor). In a recent corpus study based on spoken

language, all pronouns were collected that agreed with neuter nouns referring to persons (Strauss 2007). Of the 152 occurrences found, more than half (55%) agreed semantically rather than syntactically. This figure is a mean of very disparate percentages. Split according to target sort, the distribution of semantic and syntactic agreement is as follows (data from Strauss 2007).

(15) Semantic and syntactic agreement in spoken German

	Relative pronouns	Personal pronouns and anaphoric demonstrative pronouns	Possessive pronouns
Total number of occurrences	33	107	12
Syntactic agreement	91 %	33 %	25 %
Semantic agreement	9 %	67 %	75 %

The data, as far as it is representative, shows two interesting facts. First, with more than two thirds of the cases, semantic agreement is the norm rather than the exception everywhere but for the relative pronoun. Second, the decrease of syntactic agreement and the increase of semantic agreement correspond to the Agreement Hierarchy in the same way as was observed for Dutch.

As regards inanimate referents, German has some traces of semantic agreement, too. Two constructed examples are given in (16).[8]

(16) a) **Kartoffelsuppe**? Ja, **das** esse ich gern.
potato_soup(F) yes DEM.N eat I with_pleasure
'Potato soup? Yes, I like to eat that'

b) Sie mag keinen **Jazz**, aber ihr Freund hört
she likes no.M.SG jazz(M) but her.M friend(M) listens

das immer.
DEM.N always

'She doesn't like jazz, but her boyfriend listens to it all the time'

Here, neuter pronouns appear with feminine or masculine antecedents (though the respective feminine or masculine pronouns are also fine). Such cases are similar to the Dutch usage of the neuter, and as in Dutch, they are restricted to non-individuated referents. If the same nouns appear in a context where they refer to more individuated entities, the neuter becomes odd.

---

[8] As I am not aware of any systematic investigation of this usage, examples and judgements are based on personal judgement, corroborated by other native speakers.

(17) a) Die **Kartoffelsuppe** von gestern? **Die/?das** habe ich in den
DEF.F potato_soup(F) of yesterday DEM.F/N have I in 3.M

Kühlschrank gestellt.
fridge(M) put

'Yesterday's potato soup? I put it in the fridge.'

b) Dort wurde ganz anderer **Jazz** gespielt. **Der/?das**
there was very different.M jazz(M) played. DEM.M/N

war viel freier.
was much freer

'There they played a very different jazz. It was much freer.'

Thus, even in a thriving three-gender system, there can be some degree of semantic agreement, coupled to conceptual distinctions such as natural gender and individuation.

German differs in an interesting way from Dutch with regard to the behaviour of the possessive pronoun. While German accepts semantic and syntactic agreement on all agreement targets including the possessive, Dutch disallows syntactic agreement on possessive pronouns when this clashes with the semantics of the referent. Thus, (18a) is fine in German, while the parallel (18b) is considered ungrammatical in Dutch (ANS § 5·1·2).

(18) a) Das **Mädchen** fuhr auf **seinem** Fahrrad
DEM.N girl(N) rode on POSS.N bike

b) Het **meisje** reed op ***zijn** fiets
DEF.N girl(N) rode on POSS.N bike

'The girl rode on her bike'

Note that both in Dutch and German, the neuter possessive is syncretic with the masculine. For Dutch, one sometimes hears the argument that this makes the neuter possessive too similar to the masculine to be used comfortably for a female referent. It is interesting to consider that speakers of German, which has the same pattern of syncretism, are not hindered by such associations and freely use neuter possessives with neuter nouns referring to female persons. Apparently, Dutch personal pronouns have much stronger semantic associations than their German counterparts.

German also uses neuter pronouns in reference to events and propositions.

(19) a) Ich musste gestern   zum   Zahnarzt. **Das**   hat
        I   must   yesterday to_the.M dentist(M) DEM.N has

        eine   Stunde gedauert.
        an.F   hour(F) taken

        'I had to go to the dentist yesterday. That took an hour.'

    b) Du wolltest mich besuchen. - **Das**   habe ich nie   gesagt.
       you wanted   me   visit   DEM.N have I   never said
       'You wanted to come and see me. - I never said that.'

Events and propositions are generally expressed by verbs or clauses rather than by nouns, yet they can act as antecedents to anaphoric pronouns. In their verb-like semantics, they may be viewed as the least individuated of referents. In this light, such agreement may be yet another instance of low individuation being associated with neuter gender.

A similar idea is voiced in Enger (2004) in a discussion of the so-called Scandinavian "pancake-sentences". This term refers to non-neuter nouns taking neuter agreement on the predicative adjective, as in the two typical examples in (20) (after Enger 2004: 7).

(20) a) **Pannekaker** er **god-t**
        pancakes.PL   is good-SG.N
        'Pancakes are good'

    b) **Vodka**   er **sunt**
       vodka(M)   is healthy.N
       'Vodka is healthy'

The antecedents for which this type of agreement is found are typically indefinite nouns, mass nouns, abstracts or propositions. What unites all instances, according to Enger, is a low degree of individuation.

A similar case is reported by Hjalmar Petersen (personal communication) for Faroese.

(21) **Ostur**   er **sunnur**/**sunt** tí   **hann/*tað** innihledur calcium
     Cheese(M) is healthy.M/N   because 3.M/N   contains   calcium
     'Cheese is healthy because it contains a lot of calcium'

The predicative adjective can have masculine or neuter gender when agreeing with the masculine noun *ostur* 'cheese', just as in the Norwegian "pancake-sentences" above. Interestingly, the anaphoric pronoun in the same sentence must be masculine

(when it has a masculine antecedent). This is unexpected from the point of view of the Agreement Hierarchy, which predicts that if predicates can agree semantically, then so can personal pronouns. Norwegian is better behaved, since it also allows neuter semantic agreement on the pronoun. Enger (2004: 9) cites the following example:

(22) **Vodka**, **det**   drikker Ivan
     vodka(M) PRO.3.N drinks   Ivan
     'It's vodka that Ivan is drinking'

All of the cited cases suggest that neuter agreement with non-neuter antecedents of low individuation occurs across Germanic even without syntactic necessity, i.e. when a syntactically agreeing non-neuter form is available.

For German, mismatching pronouns in reference to inanimate entities are always optional. This distinguishes the German situation from the Dutch, in which syntactic agreement is often either not available - due to reduced paradigms - or felt to be unappropriate on semantic grounds.

9.6.2 Flemish and Brabantian

Very interesting developments can be seen in the south of the Dutch language area, i.e. in Flemish and Brabantian. These varieties still have a functioning three-gender system with three articles and three adjectival forms. The paradigms look as follows (after De Vogelaer 2006: 92).

(22) Gender agreement in Southern Dutch

Agreement target (category)	Indefinite article	Definite article	Adjective	Relative pronoun	Personal pronoun
Example forms	*nen* [M]	*den* [C]	*mooien* [M]	*die(n)* [M]	*hij* [M]
	een [F]	*de* [F]	*mooie*[F]	*die* [F]	*zij* [F]
	een [N]	*het* [N]	*mooi* [N]	*dat* [N]	*het* [N]
Number of genders	2	3	3	3[9]	3

The distinction between masculine and feminine gender in the attributive domain is weak, especially considering that the final -*n* on definite articles and adjectives is often dropped. Yet, Belgian speakers of Dutch are still aware of the two separate genders. It is therefore expected that they choose their pronouns accordingly. However, the data shows a more complex situation.

---

[9] The form *dien* for the masculine relative pronoun occurs in some dialects, though not in all, and the -*n* is probably optional to some degree, so the three-gender distinction is weak on the relative pronouns (Gunther de Vogelaer, personal communication).

In a questionnaire study with adult dialect speakers in East and West Flanders, De Vogelaer (2006 and submitted b)) reports considerable variation in pronominalization. An example is the noun *sneeuw* 'snow', which is feminine according to the standard of southern Dutch. Yet, only half of the 86 test participants (45 persons, 52.3%) actually chose a feminine pronoun. 27 persons (31.4%) opted for a masculine pronoun, while 14 (16.3%) preferred a neuter form (De Vogelaer submitted b)). According to De Vogelaer, there are two possible explanations for the masculine pronouns. Either, they can be attributed to dialect diffusion from Brabantian, the prestige variety in the east, where *sneeuw* is a masculine noun. Yet, masculine pronouns are strongest in West Flemish, where the expansion of Brabantic should be felt less strongly than in the central or eastern dialects. Alternatively, the masculines could be due to a general tendency for feminine agreements to be replaced by masculines, as also happened some 500 years earlier in northern Dutch. If this is the case, we see a weakening of the masculine-feminine distinction along the same path as was described for the nothern standard. There does not seem to be a semantic motivation for this use. Nor should formal factors be a reason, as dropping the masculine *-n* on articles and adjectives yields a feminine form, which should raise the impression that former masculines are now feminine. Yet, the use of feminine pronouns is decreasing. De Vogelaer and De Vos (submitted) analyze the patterns as "random" and ascribe them to uncertainty. Such patterns seem to be typical for an early stage of gender reduction.

For the neuter agreements observed with non-neuter nouns, De Vogelaer suggests the influence of Standard Dutch as the triggering factor. This is another example of dialect diffusion. Yet, it does not explain the neuter pronouns for *sneeuw*, which has common gender in the standard language. Here, De Vogelaer attributes neuter agreement to semantics: snow is an unindividuated mass. Hence, he describes this change as an instance of resemanticization, parallel to the developments in the north.

Interestingly, the two types of change - dialect diffusion versus system-internal tendencies such as resemanticization or overuse of the masculine - differ in domains. Changes due to dialect diffusion affect both the attributive and the pronominal domain, whereas masculine agreement for feminine nouns and neuter agreement for non-neuter mass nouns is entirely a pronominal phenomenon. This supports the assumption that the developments are of a different nature. De Vogelaer suggests that the system-internal changes are an acquisition problem: they are a case of imperfect transmission (Labov 2007) from one generation to the next (De Vogelaer submitted a)). This hypothesis, and its implications, are taken up in Chapter 10.

De Vogelaer's findings are replicated for one of the most traditional Flemish-speaking areas, the dialect of Moerzeke. In a study combining language acquisition and apparent time research, De Vogelaer and De Vos (submitted) find the same developments, in particular some masculine pronouns for feminine nouns and vice versa, which are explained as manifestations of uncertainty, and neuter gender pronouns for non-neuter mass nouns. Moreover, in the Moerzeke data, countable abstract nouns may also switch to neuter gender in anaphoric reference. Deviation from the expected gender increased monotonically from older to younger

participants. Thus, we are witnessing the decline of the traditional system. In fact, a survey on the use of gendered pronouns by 7-year-old East Flemish children, confirms that children are indeed pushing the change: they show a stronger tendency to use neuter *het* to refer to mass nouns than adults (De Vogelaer submitted a)).

The most exciting news from the studies mentioned above is that, in the south, the masculine form *hij* for neuter count nouns is only marginally observed. Instead, speakers show a very high faithfulness to neuter gender. This suggests that the shifts in gender usage start with those genders that are harder to distinguish on the attributive. The neuter, which is more distinct, is also the stronghold of syntactic agreement on the pronoun. At a later stage, it, too, may succumb to semantic pressures, as the masculine and feminine acquire stronger ties with semantics and allocate the residue to the neuter.

Thus, the reduction of the attributives and the reorganization of the pronouns do not follow each other in time, but go hand in hand. As the distinctions are weakening, pronoun usage starts to vary. The same observation is made for English by Curzan (2003). Note that Afrikaans is an interesting case: while the loss of attributive gender has long since been completed, the division of labour among the pronouns is not yet settled, as masculine and neuter pronouns still alternate in reference to inanimate entities. Yet, Afrikaans does not contradict the hypothesis that pronominal gender agreement needs to be semantic: there are no reasons to believe that the language still has syntactic gender agreement anywhere.

9.7 Summary

The cross-Germanic data confirms that personal pronouns are more resistant to deflection than articles, adjectives and other agreeing elements. 7 of the 12 standard languages (Norwegian counted twice, once on each side, for its different standard varieties) have undergone gender syncretism within the NP, while all 12 have retained the original three genders in the pronominal paradigms. This leads to a synchronic situation where more than half of the Germanic languages have more pronominal than nominal genders.

The chapter looks at the use of these supernumerary pronominal genders in English, Afrikaans, Swedish, Danish, Norwegian and Frisian. In all of these languages, they turn out to be distributed on semantic grounds. This suggests that genders which are marked only on personal pronouns cannot agree syntactically but must be organized semantically (Hyothesis B). This is not self-evident. If pronouns are one agreement target amongst several, they can agree syntactically in much the same way as, say, adjectives and predicates. Yet, they do not seem to be able to do so on their own.

A brief look at semantic agreement in more conservative languages/varieties (German, Flemish, varieties of Norwegian, Faroese) reveals similar agreement patterns, with masculine/feminine pronouns for persons and neuter pronouns for entities of low individuation. Apparently, the described semantic associations also exist in the more traditional three-way gender systems. However, semantic

agreement is optional in these varieties. When deflection cuts into the nominal genders, the semantic associations with the genders may be promoted to a more central - or indeed the only - organizing principle in pronominalization.

The relation between the loss of agreement in the noun phrase and the redistribution of the pronouns deserves closer attention. Is it indeed the case that pronouns cannot support a syntactic, non-semantic gender on their own? In the last chapter, we will investigate this issue with the help of an explorative typology of pronominal gender languages.

Chapter 10

Pronominal Gender Systems

All of the data considered so far comes from Germanic. Yet, the implications are of a general typological nature, which calls for a wider angle of view. The final chapter of this book is devoted to the connection between the reduction of agreement morphology in the attributive domain and the redistribution of the personal pronouns. This link, often taken for granted in the relevant literature, suggests a relation between gender assignment and gender agreement that is not usually recognized.

In Chapter 9, the hypothesis was offered that genders expressed solely on personal pronouns need to be semantically organized. If this is indeed the case, we may conclude that pronouns are incapable of supporting a syntactic agreement system on their own. The best way to test this hypothesis is by a cross-linguistic investigation of languages such as English: languages that mark gender entirely on personal pronouns.

10.1 Pronominal gender languages

Purely pronominal gender systems are rare, but those that could be found provide interesting evidence for the matter at hand. In the following sections, the relevant data is presented according to macro-area, before embarking on a synthesis of the observed patterns and further conclusions.

As the scarcity of information precluded proper sampling, the search was random, following suggestions in the literature. In the following, the data is reported as found, without omissions or exclusions.

10.1.1 Indo-European

In Indo-European, there are five languages with a pronominal gender system. These are Afrikaans, English, Manx (the recently extinct Celtic language formerly spoken on the Isle of Man), Persian and Yazgulyam (both Iranian languages, the latter an endangered language spoken in Tadzhikistan). English and Afrikaans were discussed in some detail in the previous chapter. The other three languages have the following gender assignment systems.

Manx has the familiar system of three genders, commonly called masculine, feminine and neuter. For Manx, there are three pronoun forms.[1] The form *ee* (pronounced /i/) is used for female persons and optionally for female animals. The form *yh* (pronounced /ə/) is used for all other referents. In addition, there is the form *eh* (pronounced /e/) which can be used for males "when the speaker wishes to make the reference clear" (Phillips 2004: 18).

The pronouns in the language are distributed according to semantic considerations. There are two splits, one between human and non-human referents and the other between male and female persons. In some cases, there is leakage, leading to pronominalization of higher animals by 'human' pronouns. For Manx, the masculine pronoun is optional: the neuter pronoun can also be used for male humans if there is no emphasis on the sex of the person (John Phillips, personal communication).

In the Iranian (Pamir) language Yazgulyam, only the oblique forms of the singular pronouns mark gender, and only in the second and the third person. This language has two genders: masculine and feminine (Payne 1989).

(1)   Yazgulyam pronominal gender, paradigms

Pronoun Gender	Oblique singular	
	2nd person	3rd person
Masculine	*day*	*way*
Feminine	*dim*	*im*

Payne describes the use of the pronouns in the following terms: "masc. gender is associated with male humans and inanimate objects, whereas fem. gender is associated with female humans and all animals (regardless of natural gender)" (Payne 1989: 429). Thus, Yazgulyam has a semantic system with the oppositions [animal or female human] and [other].

Third, there is Persian, which has lost its gender system except that it has two forms for the third person pronoun: *u*, which translates as 'he' or 'she', and *ān*, which means 'it'. The corresponding plural forms are *išān* and *ānhā*. The distribution of these forms is predictable: *u* is used for persons (and some higher animals) and *ān* for all other referents. The plural shows the same split into persons and others, with some leakage of person referents into the inanimate domain: Mace (2003: 65) notes that *ānhā* can refer to persons and/or things. Unfortunately, the source does not specify if this holds for particular things, and if so, which. Generally speaking,

---

[1] Manx has very rudimentary gender marking outside the pronominal paradigms: prepositions have forms inflected for gender. They only distinguish two genders: feminine and other. Examples are *huggey* 'to him/it' vs. *huck* 'to her' and *woish* 'from him/it' vs. *woee* 'from her' (John Phillips, personal communication). This two-gender system is typical for (Insular) Celtic. Manx is exceptional in its use of a specific masculine gender pronoun next to the more general non-feminine pronoun.

Persian, though regarded as a genderless language, shows a split between common and neuter in the personal pronouns. The distribution of the forms is based on semantics.

10.1.2 Africa

A small number of pronominal gender systems can be found in Africa. From the Niger-Congo family, four languages qualify: Zande, spoken in the Central African Republic, the Democratic Republic of Congo and the Sudan, Defaka, spoken in Nigeria, Klao, a Kru language from Liberia and Jɔ (or Jowulu), a Mande language from Mali and Burkina Faso. Moreover, Juǀ'hoan, a Khoisan language spoken in Angola, Botswana and Namibia, is a candidate. We'll briefly look at each language in turn.

Zande has four genders, expressed on personal pronouns (although gender marking is spreading to other sentence elements). Table (2) gives the paradigm (from Claudi 1985: 90, alternative forms indicate different descriptions by different grammars).

(2) Zande, third person pronouns, paradigms

Pronoun / Gender	Singular		Plural	
	Subject	Object	Subject	Object
Masculine	*ko, kù*	*ko*	*i, yo*	*yo*
Feminine	*ri, li*	*ri,li*		
Animate	*u*	*ru, l̪u*	*ami*	*ra*
Neuter	*si, ti*	*e, he, ni*	*si, ti*	*e, ni, he, ha*

The gender distinction is extended into the plural paradigm, although here the masculine and the feminine are merged.

The pattern of distribution is semantic, with a masculine gender for male persons, a feminine gender for female persons, an animate gender for animals and a residue gender for all other referents. There is occasional leakage: for small children, the animate pronoun is used, and some inanimates may take the animate pronoun. Zande thus makes a triple split, one between animates and inanimates, one between persons and animals and a sex-based split among the humans, the latter only in the singular.

Defaka (also known as Afakani) is an Ijoid language with a three-gender system that is exceptional in its genetic area, the South-Central branch of Niger-Congo. Together with its close relative, Ijo, it is the only language of this group that has a sex-based gender system.[2] The pronominal forms are given under (3).

---

[2] Ijo itself is another candidate for a pronominal gender system, but its only well-documented dialect has "a system of definite articles suffixed to the noun,

(3) Defaka third person singular pronouns (from Jenewari 1983: 103)

Pronoun / Gender	Third person singular
Masculine	*o*
Feminine	*á*
Neuter	*ye*

Despite its unusualness in this group of languages, the system is very common in terms of gender assignment: it shows the same organization as English and Manx. The pronouns *o* and *á* are used for male and female humans respectively, while *ye* refers to all other referents.

In the Mande branch of Niger-Congo, which is otherwise genderless, we find the language Jɔ, described by Carlson (1993) who explicitly compares it with Defaka. This language has a gender system that is similar to English.

(4) Jɔ personal pronouns, paradigms

Pronoun / Gender	Singular	Plural
Masculine	*ú*	*kí*
Feminine	*ní*	*kí*
Neuter	*ǹ*	*yìrì*

The three singular pronouns are distributed according to natural gender, with a threepartite split into male, female and other. The plural pronouns collapse the male/female distinction and distinguish human and other referents.

Klao, described as "Kru" by Rickard (1970), the fourth Niger-Congo language with a pronominal gender system, has an intriguing pronominal paradigm which codes not only gender and number but also verbal aspect (completive/incompletive). Gender is only distinguished in the third person singular. There are two forms, *ɔ*, translated as 'he/she' and *e* or *ɛ* (the former variant is given in Rickard 1970, the latter in Marchese 1988: 330), translated as 'it'. The forms are reduplicated in incompletive aspect, and they are identical for subject and object (first and second person pronouns do distinguish case). Unfortunately, the grammar does not explicitly give the distribution of the forms, but according to Marchese (1988: 330), the distinction is between human and non-human.

---

apparently developed from demonstratives" and among them "[a] separate feminine [...] demonstrative and definite article**"** (Williamson 1969: 6). Thus, it also has attributive gender marking and does not qualify for the type of language investigated here.

Williamson and Blench (2000) further mention a feminine in the second and third person singular pronoun in the Kru languages Niaboua (also known as Nyambwa) and Wobé (also known as Wè). Since the Kru languages are not typical gender languages, these genders might be purely pronominal; unfortunately, no grammars were available.

Another candidate among the African languages is the Khoisan language Ju|'hoan, as described by Dickens (2005) and discussed in Güldemann (2000). Güldemann explicitly analyzes it as a pronominal gender language (Güldemann 2000: 7). This language has five genders, distinguished in the pronominal paradigms. Syncretism is pervasive. Only gender V has unique forms across the different pronoun types, and gender II is *non-autonomous* (term from Baerman, Brown and Corbett 2005: 15 following Zaliznjak 1973 [2002]). This means that this gender does not have any unique markers: all of its morphological forms also occur in other genders. Table (5) gives the paradigms.

(5) Ju|'hoan pronominal genders (after Güldemann 2000)

Pronoun / Gender	Free pronoun			Pronoun as possessum		Proximal demonstrative
	Sg	Du	Pl	Sg	Pl	
I	*ha*	*sá*	*sì*	*mà*	*hì-sì*	*hè*
II	*ha*	-	*hì*	*mà*	*hì-sì*	*hè*
III	*ha*	-	*ha*	*mà*	*mà-sì*	*hè*
IV	*hì*	-	*hì*	*hì*	*hì-sì*	*hè*
V	*ká*	-	*ká*	*gá*	*gàsì*	*kè*

According to Güldemann, pronouns are distributed according to semantic and formal considerations. On the basis of field notes by Dickens (published later as Dickens 2005), Güldemann (2000) lists the following semantic classes.

> gender I: humans
> gender II: animals, other nationalities
> gender III: plants, plant food
> gender IV: long objects
> gender V: body parts, 'thing, matter', verbal nouns

However, Dickens (2005: 31) himself notes that the semantic classes are "general guide-lines" rather than accurate predictors, and there are numerous exceptions.

In my view, there is reason to doubt the status of the language as a purely pronominal gender language, i.e. as a language where gender marking is restricted to pronominal targets. Both the possessive pronouns and the free personal pronouns can be used attributively. Dickens (2005: 63) gives two examples:

(6)   ha   dshàú
        PRO.SG.I woman
        'that (previously mentioned) woman'

        ká   !aíhn
        PRO.SG.V   tree
        'that (previously mentioned) tree'

In general, Ju|'hoan has an "unusual system" (Baerman, Brown and Corbett 2005: 90), and the available grammar sketch does not answer all the questions.

10.1.3 America

Moving on to the Americas, we find pronominal gender systems in the south of the continent, particularly among the Amazonian languages. Candidates are the Mura language Pirahã, a number of Cariban languages, Parintintín and Kayabí from the Tupí-Guaraní genus, and possibly four or five of the Macro-Ge languages. Beyond the Amazon, there is the Oto-Manguean language Mixtec (Chalcatongo variety) which possibly has a relevant sort of gender system. In North America, where gender languages are scarce, interesting patterns can be found in Southeastern and Eastern Pomo. We will look at the gender assignment systems briefly.

Pirahã has three bound pronouns which can be regarded as marking gender (their free form counterpart has a single form only and thus cannot mark agreement). These forms are *xi*, *xís* and *hi* (Everett 1986) and their distribution can be schematized as follows. The feminine pronoun *xi* refers to female persons, the masculine pronoun *xís* to male humans as well as all animals, and the neuter *hi* is a residue category for inanimate referents. This system resembles many other systems of pronoun distibution we have seen, but it is unique in this particular patterning of semantic classes and genders.

Among the Cariban languages, there are several pronominal gender languages. We will only look at Hixkaryana, a language spoken in Brazil. In this language, the third-person pronouns have deictic variants with a three-way division in proximal, medial and distal. The gender split pervades through all the third-person forms.

(7) Hixkaryana third person singular pronouns (after Derbyshire 1979: 127)

Pronoun	gender I	gender II
Non-deictic	*noro*	*ɨro*
Proximal	*mosonɨ*	*onɨ*
Medial	*mokro*	*moro*
Distal	*mokɨ*	*monɨ*

The pronouns, as in all the relevant Cariban languages, are distributed according to a simple semantic pattern: the gender I set is used for animates, the gender II set for

inanimates. The other related languages are Carib, De'kwana, Panare, Tiriyo, WaiWai and Wayana (Derbyshire and Payne 1990: 54).

Tupí-Guaraní is a language group that generally lacks gender as a morphosyntactic feature, but individual languages may display gender distinctions in the paradigms of the personal pronouns. For example, the languages Parintintín and Kayabí each have three bound third-person forms (Jensen 1999).

(8) Pronominal gender in Parintintín and Kayabí

Language / Gender	Parintintín	Kayabí
masculine	*ga*	*ʔŋa*
feminine	*hẽ*	*ẽẽ*
neuter	*i-/t-/ts-*[3]	*i-/t-/ts-*

The distribution of these pronominal forms is sex-based, as in English and many other languages. Thus, we see the familiar split of [male], [female] and [other].

The same split can be observed for a number of Macro-Ge languages. Wiesemann (1989) mentions pronominal gender in the third-person pronoun paradigms of Kaingáng, Xokléng and Rikbaktsá. Besides, she notes a human/non-human distinction in the third person pronouns of Xerenté and Kayapó, which - contrary to her analysis - could be analyzed as genders. Unfortunately, no information is available on the presence or absence of gender marking in the rest of the grammar of these languages, so the issue is an open one until more material can be provided.

Outside the Amazon, a potentially interesting language is Chalcatongo Mixtec, spoken in Mexico and described by Macaulay (1996). This language has a number of third-person pronouns that still bear clear traces of their origins as nominal classifiers: they are formally identical to nouns meaning 'man', 'woman', 'animal' and so on. The gender system is not fully formed: the (clitic) pronouns are optional and there is no form for inanimate referents. Besides, there is interaction with a honorifics system. As with all classifier systems, the classes are semantically organized. Since the forms are not clearly a case of gender agreement, they should not be considered as evidence.

Finally, pronominal gender systems can be found in Eastern and Southeastern Pomo (there seem to be no gender distinctions in Central Pomo, see Mithun 1990). In both varieties, there are two gender-specific suffixes that only occur with personal pronouns of the third person (in Eastern Pomo, an exception are the nouns *qawéli-p/qawéli-t* 'second or third persons grandson/granddaughter' that also bear this marker, McLendon 1975: 164-165). In Eastern Pomo, the third person pronouns are *mí-p* for masculine and *mí-t* for feminine antecedents (McLendon 1975: 107). The

---

[3] The three neuter affixes are distributed according to stem class (Jensen 1999: 148).

paradigm for Southeastern Pomo is given in (9) (data from Moshinsky 1974). The grammar mentions a distinction between displaced and non-displaced referents, meaning "the presence or absence of the referent of the demonstrative in the speech situation. It could alternatively be termed 'visibility', since presence in the visual field is also part of the distinction" (Moshinsky 1974: 99). The category non-displaced has a three-way split according to deictic proximity.

(9) Southeastern Pomo third person singular personal pronoun, subject forms

Pronoun	Masculine	Feminine
Near	*míyi*	*mémed*
Unmarked position	*ʔúyi*	*ʔómed*
Far	*ʔíyi*	[form not mentioned in grammar]
Displaced	*yíwi*	*yímed*

Unfortunately, the sources do not disclose the exact distribution of the gender-marked pronouns. There is clearly some correlation with the natural gender of the (animate) referent, but it is uncertain if anaphoric pronouns can refer to inanimate referents and what form is used in that case. There are no other gender distinctions elsewhere in Eastern and Southeastern Pomo, making the system entirely pronominal.

10.1.4 Pronominal gender systems in Asia, Australia, New-Guinea

For the rest of the world, seven pronominal gender languages were found, though the list is certainly not exhaustive. In Australia, there are three relevant cases, two of them from different genetic groups. The first is the nearly extinct language Diyari, a Pama-Nyungan language, which has two pronominal genders. The assignment system reveals a pattern that Corbett (1991: 11) calls "noteworthy": one pronoun is used for female persons, the other for all other referents. Table (10) shows that the distinction pervades through all case variants of the third person singular personal pronoun (data from Austin 1981).[4]

(10) Diyari third person singular pronouns, paradigm

Case Gender	Ergative	Nominative	Accusative	Dative	Locative/ Allative	Ablative
Feminine	*ṉandu*	*ṉani*	*ṉaṉa*	*ṉaŋkaṉi*	*ṉaŋkaŋu*	*ṉaŋkaŋundu*
Non-feminine	*ṉulu*	*ṉawu*	*ṉiṉa*	*ṉuŋkaṉi*	*ṉuŋkaŋu*	*ṉuŋkaŋundu*

---

[4] Note that the Diyari pronouns can also be used as determiners inside the noun phrase, so the gender system is not (or no longer) entirely pronominal.

A very similar pattern is found in the related language Pitta Pitta (Blake 1979: 193-194). This language has the masculine prefix *ɲu* and the feminine prefix *ɲa* which occur on personal pronouns throughout the three cases and the three levels of deictic proximity (near, general, far). This seems to be the only morphosyntactic exponence of gender. In Pitta Pitta, the feminine forms are used for female persons and animals, the masculine forms for all other referents. Austin (1981: 60) mentions similar gender-distinguishing pronominal bases in the related languages Yandruwandha, Ngamini and Yarluyandi.

Interestingly, the same pattern returns in the unrelated Sepik-Ramu language Awtuw from Papua New-Guinea. This language has two free third-person pronouns: *tey* (feminine) and *rey* (masculine) (Feldman 1986 and personal communication). As in Diyari, the feminine pronoun refers to female persons, while the masculine pronoun can occur with any other referent.

Moving on to the next macro-area, Austro-Asiatic, we find the same gender-assignment pattern returning in the Mon-Khmer language Khmu, spoken in Laos.

(11) Khmu pronominal genders (Premsrirat 1987: 33)

Pronoun / Gender	Singular	
	Second person	Third person
Masculine	*jɛ̀ʔ*	*kə̀*
Feminine	*pà:*	*nà:*

Premsrirat lists *kə̀* as a form for masculine and neuter, but the distribution of these forms shows that the analysis in terms of a third gender is entirely based on semantic considerations, not on morphology: "masculine and feminine natural gender for humans contrast for the second and third person singular pronoun. The third person pronoun for an inanimate referent has the same form of pronoun as that of the masculine" (Premsrirat 1987: 32). Thus, Khmu resembles Diyari, Pitta Pitta and Awtuw in that it has a two-gender system employing the semantic parameters [female human] and [other].

Moreover, there is a pronominal gender language in the Sino-Tibetan family: both the Classical and the Lhasa variant of Tibetan have two personal pronouns, *kho* 'he' and *mo* 'she', which are used for male and female persons, respectively. Otherwise, the distal demonstrative *de* 'that' is used (DeLancey 2003a and b; thanks to Seth Cable for pointing out these facts). The two forms *kho* and *mo* occasionally appear on nouns as well as adjectives and seem to be a remnant of an earlier agreement system that is now lost (DeLancey 2003b: 276) and is only retained in the personal pronouns. In terms of distribution, Tibetan resembles English with a three-way split into male, female and other referents.

We conclude our world tour of pronominal gender languages with the Dravidian language Malayalam. From the description by Asher and Kumari (1997), it is clear that gender agreement in this language is practically restricted to pronouns. "The justification of noun classes [...] depends neither on morphological features [...], nor on features of concord within a noun phrase. Nor is concord with verbs relevant [...]. Gender does determine some aspects of grammar, however, and these have to do with a sort of concord. Firstly, the choice of pronoun among *avan*, *ava* and *atə* is determined by whether a noun is masculine, feminine or neuter." (Asher and Kumari 1997: 252).[5] Thus, the Malayalam personal pronouns agree in gender. The three forms are distributed according to semantic criteria. "Masculine nouns denote male human beings, feminine nouns denote female human beings, and neuter nouns nonhumans. Gods and demons are grouped with humans. Infants fall into the neuter class" (Asher and Kumari 1997: 252). In the plural, masculine and feminine are merged into a human class. Schema (12) gives the paradigm.

(12) Malayalam third person pronouns, paradigm

Pronoun / Gender	Singular	Plural
Masculine	*avan*	*avar*
Feminine	*ava*	
Neuter	*atə*	*ava*

The distibution of the singular pronouns is roughly as in English, Manx and Jɔ.

After this sketch of pronominal gender languages, it is time to return to the original question, which is whether there are commonalities among those languages with pronominal gender agreement. In particular, we wanted to know whether purely pronominal marking always coincides with semantic rather than syntactic agreement.

10.2 Synthesis: towards a typology of pronominal gender languages

If this survey is representative, there are clear patterns in the assignment systems of pronominal gender languages.

---

[5] Asher and Kumari further note that "predicative 'adjectives' (which in terms of their morphology are usually nominal in structure) must often match the subject noun in this respect" (ibid.). In my view, the latter structures do not represent an instance of agreement, as the denominal adjectives all bear the same adjectival suffix, with the gender-specificity being part of the meaning of the base noun. This seems to be comparable to forms such as *womanly* in English, which carry gender-related meaning but are not regarded as gender-marked.

The first generalization is that pronominal gender systems have a small range of genders. The largest system found is that of Zande, with four genders. However, of the 112 gender languages listed in the World Atlas of Language Structures (WALS, Haspelmath et al. 2008: 126 ff), 88 languages have systems with two to four genders. If larger systems are much less common anyway, the small group of pronominal gender languages might just mirror the general trend.

A more significant generalization is that all pronominal gender systems are semantically organized. This is not the general typological pattern: more than half of the 112 gender languages in the WALS sample have partly form-based gender assignment with phonological and/or morphological rules. Among the pronominal gender languages, only the problematic candidate Ju|'hoan has an assignment system that is not purely semantic.

Third, the assignment systems are similar in that they all make use of common, basic distinctions such as male/female, human/non-human, animate/inanimate and count/mass. These property pairs figure largely in typological research, and many grammatical distinctions are sensitive to them. Again, Ju|'hoan is the one language that illustrates the alternative: a gender system with more specific semantic classes such as long objects or body parts. Among the languages with attributive agreement, there are more of this type: Dyirbal famously has semantic classes for dangerous things (gender II) and non-flesh food (gender III) (Dixon 1972)[6], German pairs feminine gender with exotic fruit (Köpcke and Zubin 1984 and elsewhere), and in Norwegian, nouns referring to dairy products are masculine (Enger in press). Such comparably small classes are not generally found in pronominal gender languages.

A fourth generalization is that the pronominal gender languages employ the relevant semantic distinctions in such a way that the distribution of the pronouns can be aligned to a conceptual scale. In this book, the Individuation Hierarchy was suggested as a useful tool. This hierarchy sorts referents according to their degree of individuation, from sex-differentiated humans as the most individuated class to uncountable and abstract nouns as the least individuated (for the Dutch data in the preceding chapters, a more detailed version was used, containing the intermediate class *specific mass*, see Chapter 4 for motivation).

(13)

| male human
female human | > | animal | > | inanimate
object | > | mass/
abstract |

In nearly all the languages presented, the pronominal genders can be aligned to the hierarchy in such a way that each pronoun is associated with a single uninterrupted

---

[6] However, recent work by Polinsky and Plaster (in press) re-analyzes the Dyirbal system as involving more straightforward semantic rules coupled with formal gender cues. In this light, Dyirbal no longer represents a strong case for exotic semantic rules in gender assignment.

domain. The only difference between languages are the cut-off points between one gender and the next. Table (14) illustrates the situation.

Table (14) The Individuation Hierarchy and the pronominal genders[7]

Hierarchy / Language	male human / female human	>	animal	>	inanimate object	>	mass/ abstract
English, Defaka, Jɔ (SG), Parintintín, Kayabí, Tibetan, Malayalam (SG)	■ / ○		▲		▲		▲
Manx	■/▲ / ○		▲		▲		▲
Afrikaans	■ / ○		■/▲		■/▲		■/▲
Yazgulyam	■ / ○		○		■		■
Zande (SG)	■ / ○		▲		◊		◊
Persian, Jɔ (PL), Klao, Malayalam (PL)	○ / ○		■		■		■
Pirahã	■ / ○		■		▲		▲
Hixkaryana	○ / ○		○		■		■
Diyari, Pitta Pitta, Awtuw, Khmu	■ / ○		■		■		■
Zande (PL)	○ / ○		■		▲		▲

Thus, 18 of the 20 languages behave as predicted by the *Semantic Map Connectivity Hypothesis* (Haspelmath 1997, Croft 2001, 2004), which states that each element should be functionally aligned to one particular area on the scale. Besides Ju|'hoan, Yazgulyam is an exception to this principle by having a masculine gender that is used for male persons and inanimate objects. Its distributional field is interrupted by the feminine gender pronoun that is used not only for female persons, but also for all animals.

---

[7] The same symbols represent the same gender (within an individual language). When the distributions differ for singular and plural pronouns in a particular language, both variants are given. A slash indicates variation or choice between two genders.

The Khoi-San language Ju|'hoan, whose status as a pronominal gender language is contestable, violates all four generalizations. It has five genders and it is only partly semantically organized. It employs semantic rules that are cognitively less basic than those of the other systems, and these cannot be usefully aligned to a conceptual scale. In this light, more data from this language would be most welcome.

10.3 Gender assignment and gender agreement

Assuming that the outcome of this survey captures the typical traits of pronominal languages, we see that the hypothesis raised for Germanic is confirmed by the evidence from languages across the world. Purely pronominal marking indeed coincides with semantic agreement. While there are many languages whose gender system is based on phonological or morphological rules (or both), we do not find them among the pronominal gender languages. Apparently, pronouns can be the sole exponents of gender system, but only when it is simple, basic and semantically organized. For any other type, i.e. for a more elaborate or form-based system, more agreement seems to be necessary. As Curzan puts it "personal pronouns cannot indefinitely uphold a grammatical category/property-based distinction alone" (Curzan 2003: 63 after Howe 1996).

In fact, this is exactly what we have witnessed in Germanic. When genders lost their attributive agreement support - as happenend with the masculine and the feminine in Dutch, Scandinavian and Frisian, as well as with all three genders in English and Afrikaans - speakers could no longer reliably tell them apart. The availability of gender-marked personal pronouns could not prevent this development. Instead, the genders in question were redistributed according to meaning (resemanticization).

What is it, then, that makes pronouns such weak representatives for a gender system? The issue is probably related to acquisition. In Chapter 9, we saw that the changes in Flemish pronoun usage were attributed to incomplete transmission (De Vogelaer submitted a)). If this holds true, then attributive gender markers are necessary for non-semantic gender distinctions to be acquired. The personal pronouns, by virtue of their distance to the noun as well as their relative semantic and syntactic freedom, are poor cues for gender acquisition. Moreover, pronominal genders lack the redundancy of repeated marking that is found in more canonical agreement systems, where gender is marked on several exponents (say, articles, adjectives and predicates). For the acquisition of a gender system with complex rules, repeated marking may be essential.

Theoretically, the facts from pronominal gender systems are a challenge to the received view on the relation between gender assignment and gender agreement. Usually, gender agreement is seen as secondary to gender assignment: agreement merely expresses the choices made during the assignment process. However, if the evidence from pronoun gender is considered, we can conclude that agreement can actively constrain assignment. When agreement is reduced to the pronominal domain, assignment rules must readjust themselves and genders are restructured along semantic terms. Evidence can be found in pronominal gender languages

213

throughout the world, as well as in languages where individual genders are expressed pronominally. All of these systems are semantically organized to begin with or develop semantic rules in the course of time.

10.4 Summary

In the last chapter of this book, the facts from Dutch gender agreement are linked to the wider typological picture. In search of evidence for a link between pronominal gender marking and semantic agreement, we looked at pronominal gender languages throughout the world. For 20 languages of this type, the available literature provided data on the distribution of the genders. The results were in line with the facts from Germanic in general and Dutch in particular. Cross-linguistically, genders only marked pronominally are based on semantic distinctions. These distinctions are cognitively general and basic, and they employ property pairs such as male/female, human/animal, animate/inanimate or count/mass. While these represent the semantic core of gender systems in general, the lack of other, notably formal rules in pronominal gender languages is significant. The facts suggest that genders relying on pronominal exponence must be semantics-based.

Turning the argument around, there are good reasons to believe that phonological and morphological gender assignment rules, as well as more complex rules in general, need more formal support than the pronouns can provide. This is interesting for linguistic theory, as it suggests that gender assignment can be constrained by gender agreement. For Dutch, it explains why the merger of masculine and feminine attributive agreements resulted in the loss of the distinction in the speakers' grammar. The pronouns retained masculine and feminine markers, but were unable to uphold the original genders as they were. Instead, the syntactic system was replaced by a semantic one which now governs pronoun usage in the spoken language. Thus, the quirky Dutch facts fit well with the general typological patterns.

Conclusion

This book presents an in-depth investigation of the pronominal gender system of modern spoken Dutch. The language exhibits an interesting paradigmatic mismatch: there are two genders marked on the definite article, the adjective, and the relative pronoun, while the personal pronouns distinguish four different genders. This contradicts our expectations about agreement systems: the controller and the targets should have the same feature values. For Dutch, the mismatch goes hand in hand with a bewildering variation in pronominal gender agreement. Common gender nouns can take masculine, feminine or neuter agreements, while neuter nouns appear with masculine or feminine personal pronouns or common gender demonstratives. Again, this is counter to the normal expectations. After all, a noun should consistently trigger the same gender on its agreeing elements.

The mismatch situation has its roots in the language history. When the masculine and feminine gender markers became indistinguishable in the noun phrase, speakers lost the knowledge about the gender affiliation of the non-neuter nouns. Official grammar writing and lexicography attempted to conserve the original system and devised word lists, where the gender of nouns could be looked up for the sake of correct pronominalization. Although this policy has been relaxed in recent times, there is still a considerable gap between spoken and written language. The aim of this study was to find out how the spoken language, which literally has no time for dictionaries, has solved the pronominal problem. Therefore, a large sample of spontaneous speech from the Corpus of Spoken Dutch was investigated in search for the pronominalization strategies in modern colloquial Dutch.

The theoretical approach taken is that the Dutch pronouns are agreeing elements in much the same way as the articles or the adjectives. Thus, pronouns with the 'wrong' gender are not set aside as non-agreeing, but rather seen as agreeing with different properties of the noun. From what is known about morphosyntactic features, these properties are expected to be semantic.

A corpus study of a 500,000 word sample of colloquial speech shows that, indeed, Dutch speakers choose their pronouns on the basis of semantic patterns. Masculine pronouns are used for male persons, for all animals (even for animals of female sex) as well as for countable, bounded objects and abstracts. Neuter pronouns, by contrast, appear in combination with mass nouns and uncountable, unbounded, unspecific abstracts. Feminine pronouns have the most restricted distribution: they can only refer to female persons and (occasionally) female animals. Whenever a

pronoun diverged in gender from its antecedent noun, these semantic rules could be seen in operation.

It is argued that these usage patterns can be combined into a unified account for semantics-based pronoun choice in spoken Dutch. Pronominalization is sensitive to the ***degree of individuation*** of the referent. Highest in individuation are persons, followed by animals and objects. This is the domain of the masculine (female persons and animals aside; for those, the feminine is used). The least individuated referents are masses and unbounded abstracts, which are associated with neuter gender. In between lies the class of specific masses, which combines properties of objects and properties of substances. Here, the domains of the masculine and the neuter meet.

The degree of individuation of a referent largely depends on the construal of situations and their participants. The noun *lamb* can refer to a highly individuated pet, an unspecific animal within a flock or even the ingredient of a meal. In Dutch, pronominalization is sensitive to such differences. This explains that the same noun can appear with various pronominal genders under different circumstances. Such variation can occur within the speech of an individual speaker or between speakers. Even in written texts, examples can be found, although the written standard dictates syntactic agreement between anaphors and their antecedent nouns.

Explained in semantic terms, the seemingly chaotic pronoun use in spoken Dutch is shown to be systematic and regular. Dialect studies indicate that gender agreement based on individuation occurs in other varieties of Indo-European and beyond. Thus, the Dutch facts are neither isolated nor exotic.

Next to the semantic system described in this study, speakers also employ the traditional syntactic gender system. This system pairs neuter nouns with neuter pronouns and non-neuter nouns with the demonstratives *deze* or *die*. The choice between syntactic and semantic agreement is influenced by a variety of factors. Statistical analysis of the data shows that semantic agreement is more likely

- for personal than for relative pronouns (in line with the Agreement Hierarchy),
- for full form pronouns than for clitics or demonstrative pronouns
- for nominative than for oblique pronouns
- for nouns at the extreme ends of the Individuation Hierarchy than for nouns in the middle

Also, the likelihood for semantic agreement could be seen to increase with a greater distance between the noun and the pronoun. 'Switching back' after a semantically agreeing pronoun has been chosen is rare. Moreover, semantic agreement was shown to be the more progressive option: speakers above 60 years of age use it only half as much as speakers below 20. This suggests that the semantic system is on the rise.

With the help of the Individuation Hierarchy and the Agreement Hierarchy much of the variation can be accounted for, either in terms of conceptual construal, or as a competition between two alternative agreement systems.

The Dutch pronominal problem and its solution is placed in a cross-Germanic context. Perhaps surprisingly, more than half of the standard languages, i.e. English, Afrikaans, Danish, Swedish, varieties of Norwegian, Dutch and Frisian, have more pronominal than attributive genders. For all cases, it can be shown that those genders that are marked only pronominally obey semantic rules. This holds true even if the rest of the gender system is not primarily semantics-based. This suggests that there is something special about personal pronouns: they cannot support a syntactic distinction on their own. This inability is the reason why many of the Germanic pronominal genders have developed new usage patterns on semantic grounds.

The assumed causality between loss of attributive agreement markers and reorganization of the pronouns leads to the hypothesis that some types of agreement are better support for a gender system than others. In order to see if this is cross-linguistically valid, an explorative typology of pronominal gender systems was conducted. 20 relevant languages were found. Nearly all of them employ simple, basic, cross-linguistically common semantic rules. This invites the conclusion that personal pronouns can only support strictly semantic systems based on general, cognitively basic assignment rules. Apparently, more complex gender systems need more support by repeated marking, preferably in the local domain of the noun phrase.

In this light, the unusual Dutch facts tie in well with the general typology of pronominal gender languages. This confirms the naturalness of the development from syntactically agreeing to increasingly semantically agreeing pronominal genders.

When regarded with unbiased eyes, the 'wrong' pronouns of spoken Dutch represent a useful and ingenious case of recovery from a historical problem. Speakers of Dutch have 'reinvented' their pronoun genders by putting new semantic foundations under the gender system.

References

Allan, Keith. 1980. Nouns and Countability. *Language* 56:541-567.

Allan, Robin, Holmes, Philip, and Lundskær-Nielsen, Tom. 2000. *Danish: An essential grammar*. London: Routledge.

Anderson, Lloyd. 1982. The 'Perfect' as a Universal and as a Language-Particular Category. In *Tense - Aspect: Between Semantics and Pragmatics*, ed. Paul J. Hopper, 227-264. Amsterdam: Benjamins.

Ariel, Mira. 1996. Referring Expressions and the +/- Coreference Distinction. In *Reference and Referent Accessibility*, eds. Thorstein Fretheim and Jeanette K. Gundel, 13-35. Amsterdam: Benjamins.

Asher, R.E. and T.C. Kumari. 1997. *Malayalam*. London: Routledge.

Audring, Jenny. 2006a. Pronominal Gender in Spoken Dutch. *Journal of Germanic Linguistics* 18:85-116.

Audring, Jenny. 2006b. Genusverlies en de betekenis van voornaamwoorden. In *Nederlands tussen Duits en Engels*, eds. Matthias Hüning, Ulrike Vogl, Ton van der Wouden, Arie Verhagen, 71-88. Leiden: Stichting Neerlandistiek Leiden.

Audring, Jenny. to appear. Pronominales Genus. In *Kontrastive germanistische Linguistik*, eds. Damaris Nübling, Antje Dammel and Sebastian Kürschner. Hildesheim: Georg Olms Verlag.

Audring, Jenny. (2009). Gender Assignment and Gender Agreement. *Morphology* 18:93-116.

Austin, Peter. 1981. *A grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.

Baerman, Matthew, Dunstan Brown and Greville G. Corbett. 2005. *The Syntax-Morphology-Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

Baker, Mark C. 1996. *The polysynthesis parameter*. New York etc.: Oxford University Press.

Bandle, Oskar et al. 2002. *The Nordic languages : an international handbook on the history of the North Germanic languages*. Berlin: de Gruyter.

Barlow, Michael. 1991. The agreement hierarchy and grammatical theory. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistic Society*, eds. Laurel A. Sutton, Christopher Johnson and Ruth Shields, 30-40. Berkeley: Berkeley Linguistic Society.

Barlow, Michael. 1992. *A Situated Theory of Agreement*. New York, London: Garland Publishing, Inc.

Bechert, Johannes. 1982. Grammatical gender in Europe: An areal study of a linguistic category. *Papiere zur Linguistik* 26:23-33.

Blake, Barry J. 1979. Pitta Pitta. In *Handbook of Australian Languages*, eds. Robert M. W. Dixon and Barry J. Blake, 183-242. Amsterdam: John Benjamins.

Bloomfield, Leonard. 1933. *Language*. New York: Holt and Company.

Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18:355-387.

Bosch, Peter. 1983. *Agreement and Anaphora*. London: Academic Press.

Branigan, H.P., Pickering, M.J., and Cleland, A.A. 2000. Syntactic coordination in dialogue. *Cognition* 75:813-825.

Braunmüller, Kurt. 2000. Gender in North Germanic: a diasystematic and functional approach. In *Gender in Grammar and Cognition*, eds. Barbara Unterbeck and Matti Rissanen, 25-54. Berlin: Mouton de Gruyter.

Bresnan, Joan, and Mchombo, Sam A. 1986. Grammatical and anaphoric agreement. In *Proceedings of the Chicago Linguistic Society 22. Papers from the Parasession on Pragmatics and Grammatical Theory*, eds. Anne Farley, Peter Farley and Karl McCullough, 741-782. Chicago: Chicago Linguistic Society.

Bunt, Harry C. 1985. *Mass Terms and Model-Theoretic Semantics*. Cambridge: Cambridge University Press.

Bybee, Joan L. 1985. *Morphology*. Amsterdam: Benjamins.

Carlson, Robert. 1993. A sketch of Jo: a Mande language with a feminine pronoun. *Mandenkan* 25, 1-109.

Caron, W.J.H. 1953 [1625]. *De Nederduytsche grammatica ofte spraec-konst van Christiaen van Heule*. Groningen: Wolters.

Caron, W.J.H. 1958. *Aanmerkingen op de neederduitsche taale en Naaberecht door Petrus Leupenius*. Groningen: Wolters.

Chierchia, Gennaro. 1998. Plurality of mass nouns and the notion of "semantic parameter". In *Events and Grammar*, ed. Susan D. Rothstein, 53-104. Dordrecht: Kluwer.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Classen, E. 1919. On the Origin of Natural Gender in Middle English. *Modern Language Review* 14:97-102.

Claudi, Ulrike. 1985. *Zur Entstehung von Genussystemen*. Hamburg: Helmut Buske Verlag.

Comrie, Bernard. 2003. When Agreement Gets Trigger-Happy. In *Agreement: A Typological Perspective*, eds. Dunstan Brown, Greville G. Corbett and Carole Tiberius, 131-337. Oxford: Blackwell.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15:203-224.

Corbett, Greville G. 1983. *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. London: Croom Helm.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2001. *Agreement: terms and boundaries. The Role of Agreement in Natural Language - Proceedings of the 2001 Texas Linguistic Society Conference, Austin, Texas, 2-4 March 2001*. Texas.

Corbett, Greville G. 2003a. Agreement: Canonical instances and the extent of the phenomenon. In *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001)*, eds. Geert Booij, Janet DeCesaris, Angela Ralli and Sergio Scalise, 109-

128. Barcelona: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra.

Corbett, Greville G. 2003b. Agreement - The Range of the Phenomenon and the Principles of the Surrey Database of Agreement. In *Agreement: A Typological Perspective*, eds. Dunstan Brown, Greville G. Corbett and Carole Tiberius, 155-202. Oxford: Blackwell.

Corbett, Greville G. 2005. The canonical approach in typology. In *Linguistic Diversity and Language Theories*, eds. Zygmunt Frajzyngier, Adam Hodges and David S. Rood, 25-49. Amsterdam: John Benjamins.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2007. Canonical typology, suppletion and possible words. *Language* 83:8-42.

Cornish, Francis. 1986. *Anaphoric Relations in English and French. A Discourse Perspective*. London: Croom Helm.

Cornish, Francis. 1999. *Anaphora, Discourse, and Understanding*. Oxford: Clarendon Press.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Croft, William. 2004. *Typology and Universals*. Cambridge: Cambridge University Press.

Croft, William et al. 1987. Diachronic Semantic Processes in the Middle Voice. In *Papers from the 7th International Conference on Historical Linguistics*, eds. Anna Giacalone Ramat, Onofrio Carruba and Giuliano Bernini, 179-192. Amsterdam: Benjamins.

Curzan, Anne. 2003. *Gender Shifts in the History of English*. Cambridge: Cambridge University Press.

Dahl, Östen. 2000. Animacy and the notion of semantic gender. In *Gender in Grammar and Cognition*, eds. B.arbara Unterbeck and Matti Rissanen, 99-115. Berlin: Mouton de Gruyter.

Davidson, Herbert. 1990. *Han hon den*. Lund: Lund University Press.

De Bonth, Rolandus J.G., and Dibbets, Gerardus R.W. 1995. *Voor rede vatbaar: tien voorredes uit het grammaticale werk van Van Hoogstraten, Nylöe, Moonen, Sewel, Ten Kate, Huydecoper, (1700-1730)*. Münster: Nodus

Dekeyser, Xavier. 1980. The diachrony of the gender systems in English and Dutch. In *Historical Morphology*, ed. Jacek Fisiak, 97-111. The Hague: Mouton.

DeLancey, Scott. 2003a. Classical Tibetan. In *The Sino-Tibetan Languages*, eds. Graham Thurgood and Randy J. LaPolla, 255-269. London: Routledge.

DeLancey, Scott. 2003b. Lhasa Tibetan. In *The Sino-Tibetan Languages*, eds. Graham Thurgood and Randy J. LaPolla, 270-288. London: Routledge.

Demuth, Katherine, Faraclas, Nicholas, and Marchese, Lynell. 1986. Niger-Congo noun class and agreement systems in language acquisition and historical change. In *Noun Classes and Categorization*, ed. Colette G. Craig, 453-471. Amsterdam: Benjamins.

Denny, J.P., and Creider, A. 1976. The semantics of noun classes in Proto-Bantu. *Studies in African Linguistics* 7:1-30.

Derbyshire, Desmond C. and Doris L. Payne. 1990. Noun Classification Systems of Amazonian Languages. In *Amazonian Linguistics. Studies in Lowland*

South American Languages, ed. Doris L. Payne, 243-271. Austin: University of Texas Press.

Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North Holland Publishing.

Deumert, Ana. 2004. *Language Standardization and Language Change: The dynamics of Cape Dutch*. Amsterdam: John Benjamins.

Dibbets, Gerardus R.W. 1995. *De woordsoorten in de Nederlandse triviumgrammatica*. Amsterdam: Neerlandistiek VU, Münster: Nodus.

Dickens, Patrick J., ed. by Rainer Vossen and Megan Biesele. 2005. *A concise grammar of Ju'hoan*. Cologne: Köppe.

Diessel, Holger. 1999. *Demonstratives. Form, Function and Grammaticalization*. Amsterdam: Benjamins.

DiSciullo, Anna M., and Williams, E. 1987. *On the Definition of Word*. Cambridge, Mass.: MIT Press.

Dixon, R.M.W. 1972. The Dyirbal Language of North Queensland. Cambridge: Cambridge University Press.

Dixon, Richard M.W. 1979. Ergativity. *Language* 55:59-138.

Donaldson, Bruce C. 1993. *A Grammar of Afrikaans*. Berlin: Mouton de Gruyter.

Donaldson, Bruce C. 2000. *Colloquial Afrikaans*. London: Routledge.

Duke, Janet. to appear. Gender reduction and loss in Germanic: The Scandinavian, Dutch, and Afrikaans case studies. In *Kontrastive germanistische Linguistik*, eds. Damaris Nübling, Antje Dammel and Sebastian Kürschner. Hildesheim: Georg Olms Verlag.

Elworthy, Frederic T. 1879 [1965]. *Specimens of English Dialects. Devonshire and Exmoor and Courtship*: English Dialect Society 25. Vaduz: Kraus Reprint LTD.

Enger, Hans-Olav. 2004. Scandinavian pancake-sentences as semantic agreement. *Nordic Journal of Linguistics* 27:5-34.

Enger, Hans-Olav. (in press). The role of core and non-core semantic rules in gender assignment. *Lingua*.

Evans, Nicholas. 1999. Why argument affixes in polysynthetic languages are not pronouns: evidence from Bininj Gun-wok. *Sprachtypologie und Universalienforschung* 52:255-281.

Everett, Daniel L. 1986. Pirahã. In *Handbook of Amazonian Languages*, eds. Desmond C. Derbyshire and Geoffrey Pullum, 200-325. Berlin: Mouton de Gruyter.

Feldman, Harry. 1986. *A grammar of Awtuw*. Canberra: Department of Linguistics, Research School of Pacific Studies, the Australian National University.

Fernández-Ordóñez, Inés. 2006-2007. Del Cantábrico a Toledo: El "neutro de materia" hispánico en un contexto románico y tipológico. *Revista de Historia de la Lengua Española* 1. 67-118; 2. 29-81.

Fernández-Ordóñez, Inés. to appear. The development of mass/count distinctions in Indo-European languages. In *Grammatical changes in Indo-European languages*, ed. V. Bubenik, Hewson, J., Rose, S. Amsterdam/Philadelphia: John Benjamins.

Field, Andy. 2005. *Discovering Statistics Using SPSS. 2nd ed.* London: Sage.

Fletcher, William H. 1987. Semantic Factors in Dutch Gender Choice. In *Papers from the Second Interdisciplinary Conference on Netherlandic Studies held*

at Georgetown University, 7-9 June 1984, ed. William H. Fletcher, 51-63. Lanham: University Press of America.

Fokkema, Klaas. 1968. De historie van enkele pronomina in het Westlauwers Fries. In *Philologia Frisica Anno 1966*, ed. Klaas Fokkema. Grins: Wolters-Noordhoff.

Fraurud, Kari. 1996. Cognitive Ontology and NP Form. In *Reference and Referent Accessibility*, eds. Thorstein Fretheim and Jeanette K. Gundel, 65-87. Amsterdam: Benjamins.

Gachelin, Jean-Marc. 1991. Gender and Deixis in Southwestern Dialects. *Neuphilologische Mitteilungen* XCII:83-93.

Geeraerts, Dirk. 1992. Pronominale-Masculiniserings-Parameters in Vlaanderen. In *De Binnenbouw van het Nederlands: Een bundel artikelen voor Piet Paardekooper*, eds. Hans Bennis and Jan W. de Vries, 73-84. Dordrecht: ICG Publications.

Geerts, Guido. 1966. *Genus en Geslacht in de Gouden Eeuw: Een bijdrage tot de studie van de nominale klassifikatie en daarmee samenhangende adnominale flexievormen en pronominale verschijnselen in Hollands taalgebruik van de zeventiende eeuw*. Brussel: Belgisch Interuniversitair Centrum Voor Nederlandistiek.

Geerts, Guido. 1995a. Genusfouten: Hollanditis in Vlaanderen? In *Nederlands, een en veelzijdig - Festschrift voor Guido Geerts*, ed. Guido Geerts, 67-76. Leuven: Universitaire Pers.

Geerts, Guido. 1995b. Hij geeft melk. In *Nederlands, een en veelzijdig - Festschrift voor Guido Geerts*, ed. Guido Geerts, 43-52. Leuven: Universitaire Pers.

Geerts, Guido. 1995c. Is het gemeentebestuur haar boekje te buiten gegaan? In *Nederlands, een en veelzijdig - Festschrift voor Guido Geerts*, ed. Guido Geerts, 53-66. Leuven: Universitaire Pers.

Gestel, Frank Van, Twilhaar, Jan Nijen, Rinkel, Tineke, and Weerman, Fred. 1992. *Oude zinnen. Grammaticale analyse van het Nederlands tussen 1200-1700*. Leiden: Martinus Nijhoff.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Görlach, Manfred. 2001. *A Dictionary of European Anglicisms*. Oxford: Oxford University Press.

Güldemann, Tom. 2000. Noun categorization in the Non-Khoe lineages of Khoisan. *Afrikanische Arbeitspapiere Köln* 63, 5-33.

Grasserie, R. de la. 1898. La catégorie psychologique de la classification, révélée par le langage. *Revue philosophique de la France at de l'étranger* 45:594-624.

Haas, Mary R. 1940. *Tunica*. New York: J.J. Augustin.

Haase, Martin. 2000. Reorganization of a gender system: The Central Italian neuters. In *Gender in Grammar and Cognition*, eds. Barbara Unterbeck and Matti Rissanen, 221-236. Berlin: Mouton de Gruyter.

Haeringen, Coenraad B. Van. 1936. *Geslacht, verbuiging, voornaamwoordelijke aanduiding : rapport uitgebracht aan Zijne Excellentie den Minister van Onderwijs, Kunsten en Wetenschappen* 's Gravenhage: Van Goor Zonen.

Haeringen, Coenraad B. van. 1954. *Genus en Geslacht - Het voornaamwoordelijk gebruik in de gesproken taal*. Amsterdam: J.M. Meulenhoff.

Halliday, M.A.K., and Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Hämeen-Anttila, Jaako. 2000. Grammatical gender and its development in Classical Arabic. In *Gender in Grammar and Cognition*, eds. Barbara Unterbeck and Matti Rissanen, 595-608. Berlin: Mouton de Gruyter.

Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie (eds.). 2008. *WALS Online*. Munich: Max Planck Digital Library.

Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford: Oxford University Press.

Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase : zur Emergenz syntaktischer Struktur* Tübingen: Niemeyer.

Hoekstra, Jarich. 1996. Transitive Pronouns and Gender Syncretism in Fering-Öömrang (North Frisian). *NOWELE* 27, 45-66.

Hoekstra, Jarich. 2006. On the loss of feminine gender in Fering-Öömrang (North Frisian). Paper presented at *Amsterdam Gender Colloquium, 15-16 September 2006*, Amsterdam, Netherlands.

Hogenhout-Mulder, Maaike. 1983. *Cursus Middelnederlands*. Groningen: Wolters-Noordhoff.

Hogenhout-Mulder, Maaike, and Reenen, Pieter van. 1988. Buiging in de zelfstandignaamwoordsgroep. In *Uitgangspunten en toepassingen*, ed. J.A. Van Leuvensteijn, 157-173. Amsterdam: VU Uitgeverij.

Holes, Clive. 1994. *Modern Arabic: Structures, Functions, and Varieties*. London: Longman.

Holmes, Philip, and Hinchcliffe, Ian. 2003. *Swedish: A comprehensive grammar*. London: Routledge.

Hoppenbrouwers, Cor. 1990. *Het regiolect. Van dialect tot Algemeen Nederlands*. Muiderberg: Coutinho.

Hopper, Paul J., and Thompson, Sandra A. 1980. Transitivity in Grammar and Discourse. *Language* 56:251-299.

Howe, Stephen. 1996. *The Personal Pronouns in the Germanic Languages. A Study of Personal Pronoun Morphology and Change in the Germanic Languages from First Records to the Present Day*. Berlin: de Gruyter.

Huddleston, Rodney, and Pullum, Geoffrey K. 2002. *The Cambridge grammar of the English language*. Cambridge Cambridge University Press.

Ihalainen, Ossi. 1985. '*He took the bottle and put 'n in his pocket*: The object pronoun *it* in present-day Somerset', in: *Focus on: England and Wales*. ed. Wolfgang Viereck, 153-161. Amsterdam: Benjamins.

Ihalainen, Ossi. 1991. On grammatical diffusion in Somerset folk speech, In *Dialects of English: Studies in Grammatical Variation*, eds. Peter Trudgill and J.K. Chambers, 104-119. London: Longman.

Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41:9-45.

Janssen, Niels, and Caramazza, Alfonso. 2003. The selection of closed-class words in noun phrase production: The case of Dutch determiners *Journal of Memory & Language* 48:635-652.

Jenewari, Charles E.W. 1983. Defaka, Ijo's Closest Linguistic Relative. In *Current approaches to African Linguistics*, ed. Ivan R. Dihoff, 85-111. Dordrecht: Foris.

Jensen, Cheryl. 1999. Tupí-Guaraní. In *The Amazonian languages*, eds. R. M. W. Dixon and Alexandra Y. Aikhenvald, 125-63. Cambridge: Cambridge University Press.

Kathol, Andreas. 1999. Agreement and the syntax-morphology interface in HPSG. In *Studies in Contemporary Phrase Structure Grammar*, eds. Robert D. Levine and Georgia M. Green, 223-274. Cambridge: Cambridge University Press.

Keenan, Edward L. and Bernard Comrie. 1977. Noun Phrase Accessibility and Universal Grammar, *Linguistic Inquiry* 8, 63-99.

Klijnhout-Moerman, T., and E.S.F. Feteris 1982. *Schrijvenderwijs*. Amsterdam: Meulenhoff Educatief.

Kollewijn, Roeland A. 1916 *Opstellen over spelling en verbuiging*. Groningen: Wolters.

Köpcke, Klaus-Michael and David Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation, *Linguistische Berichte* 93, 26–51.

Koptjevskaja-Tamm, Maria. 2004. Mass and Collection. In *Morphology: A Handbook on Inflection and Word Formation*, eds. Geert Booij, Christian Lehmann and Joachim Mugdan, 1016-1031. Berlin: Walter de Gruyter.

Kruisinga, E. 1924. *A grammar of modern Dutch*. London: Allen & Unwin.

Kuno, Susumu, and Kaburaki, Etsuko. 1977. Empathy and Syntax. *Linguistic Inquiry* 8:627-672.

Labov, William. 2007. Transmission and diffusion. *Language* 83:344-387.

Lambrecht, Knut. 2001. Dislocation. In *Language Typology and Language Universals: An International Handbook*, eds. Matrin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible, 1050-1078. Berlin: Walter de Gruyter.

Langacker, Ronald W. 1991. *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.

Lass, Roger. 1992. Phonology and morphology, In *The Cambridge History of the English Language. Volume II, 1066-1476*, ed. Norman Blake, Cambridge: Cambridge University Press, 23-155.

Lehmann, Christian. 1988. On the Function of Agreement. In *Agreement in Natural Languages: Approaches, Theory, Descriptions*, eds. Michael Barlow and Charles A. Ferguson, 55-66. Stanford, CA: CSLI.

Leuvensteijn, J.A. van. 1986. Genus en geslacht in het dagboek van Wouter Jacobsz. *Voortgang : jaarboek voor de Neerlandistiek* 7:113-136

Leuvensteijn, J.A. van. 1992. Genus in zestiende-eeuws Nederlands. Een verkennende studie. *Acta Universitatis Wratislaviensis, Studia Neerlandica et Germanica* 1356:229-237.

Leuvensteijn, J.A. van. 1997. De genusclassificatie in het Hollandse Reisjournaal van Jan Martensz. Merens uit 1600. *Taal en tongval* 49:145-157.

Leuvensteijn, J.A. van and Dekker, M.C.H. 1990. Adnominale flexie en genusmarkering in 16de-eeuws Brugs *Leuvensche bijdragen op het gebied van de Germaansche philologie en in 't bijzonder van de Nederlandsche dialectkunde* 79:257-278

Levin, Magnus. 2001. *Agreement with Collective Nouns in English*. Stockholm: Almqvist & Wiksell.

Lyons, John. 1977. *Semantics Vol. II*. Cambridge: Cambridge University Press.

Macaulay, Monica. 1996. *A grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.

Mace, John. 2003. *Persian Grammar*. London: Routledge.

MacKay, Donald G., and Konishi, Toshi. 1980. Personification and the pronoun problem. *Women's Studies International Quarterly* 3:149-163.

Maljaars, Abraham. 1979. *Bijdrage tot de kennis van genus en geslacht in de achttiende eeuw met een uitwijding over de pronominale aanduiding in het moderne Nederlands*. Utrecht: Instituut De Vooys.

Marchese, Lynell. 1988. Noun Classes and Agreement Systems in Kru: A Historical Approach. In *Agreement in Natural Languages: Approaches, Theory, Descriptions*, eds. Michael Barlow and Charles A. Ferguson, 323-341. Stanford, CA: CSLI.

Marcoux, Dell R. 1973. Deviation in English gender. *American Speech* 48:98-107.

Matasović, Ranko. 2004. *Gender in Indo-European*. Heidelberg: Universitaetsverlag Winter.

Mathiot, Madeleine, and Marjorie, Roberts. 1979. Sex roles as revealed through referential gender in American English. In *Ethnolinguistics: Boas, Sapir and Whorf Revisited*, ed. Madeleine Mathiot, 1-47. The Hague: Mouton.

McLendon, Sally. 1975. *A Grammar of Eastern Pomo*. Berkeley: University of California Press.

Mithun, Marianne. 1990. Third-person reference and the function of pronouns in Central Pomo natural speech. *International Journal of American Linguistics* 56(3), 361-376.

Mithun, Marianne. 2003. Pronouns and agreement: the information status of pronominal affixes. In *Agreement: A Typological Perspective*, eds. Dunstan Brown, Greville G. Corbett and Carole Tiberius, 235-278. Oxford: Blackwell.

Mooijaart, Marijke, and Wal, Marijke van der. 2008. *Nederlands van Middeleeuwen tot Gouden Eeuw: cursus Middelnederlands en Vroegnieuwnederlands* Nijmegen: Vantilt.

Moore, Samuel 1921. Grammatical and natural gender in Middle English. *PMLA*, pp. 79-103.

Moravcsik, Edith. 1988. Agreement and markedness. In *Agreement in Natural Language*, eds. Michael Barlow and Charles A. Ferguson, 89-106. Stanford: CSLI.

Morsbach, Lorenz 1926. *Grammatisches und psychologisches Geschlecht im Englischen*. Berlin: Weidmannsche Buchhandlung.

Moshinsky, Julius. 1974. *A Grammar of Southeastern Pomo*. Berkeley: University of California Publications.

Paddock, Harold. 1991. The actuation problem for gender change in Wessex versus Newfoundland. In *Dialects of English*, eds. John K. Chambers and Peter Trudgill, 29-47. London, NY: Longman.

Payne, John R. 1989. Pāmir languages. In *Compendium Linguarum Iranicarum*, ed. R. Schmitt, 417-444. Wiesbaden: Reichert.

Pawley, Andrew 2002. Using he and she for inanimate nouns in English: Questions of grammar and worldview. In *Ethnosyntax: Explorations in Language and Culture*, ed. Nick Enfield, 110-137. Oxford: Oxford University Press.

Pawley, Andrew. 2004. Australian Vernacular English: Some grammatical characteristics. In *Varieties of English*, eds. Bernd Kortmann and Edgar Schneider, 611-642. Berlin: Mouton de Gruyter.

Phillips, John D. 2004. *Manx*. München: Lincom Europa.

Polinsky, Maria, and Plaster, Keith. in press. Features in categorization, or a new look at an old problem. In *Features: Perspectives on a Key Notion in Linguistics*, eds. Anna Kibort and Greville G. Corbett. Oxford: Oxford University Press.

Pollard, Carl, and Sag, Ivan A. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Ponelis, Fritz. 1979. *Afrikaanse sintaksis*. Pretoria: van Schaik.

Ponelis, Fritz. 1993. *The Development of Afrikaans*. Frankfurt(Main): Peter Lang.

Premsrirat, Suliwai. 1987. *Khmu, a minority language of Thailand*. Canberra: Australian National University.

Priestly, Tom M.S. 1983. On 'Drift' In Indo-European Gender Systems. *Journal of Indo-European Studies* 11:339-363.

Pullum, Geoffrey K., and Zwicky, Arnold M. 1988. The syntax-morphology interface. In *Linguistics. The Cambridge Survey, Vol. I*, ed. Frederick Newmeyer, 255-280. Cambridge: Cambridge University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman

Rickard, David T. 1970. *Kru grammar: a preliminary study*. Monrovia: Department of Literacy Work and Literature Production of the United Methodist Church.

Ringgaard, K. 1973. *Danske dialekter. En kortfattet oversigt*. København: Akademisk Forlag.

Rohdenburg, Günter. 2004a. Comparing grammatical variation phenomena in non-standard English and Low German dialects from a typological perspective. In *Dialectology meets Typology. Dialect Grammar from a Cross-Linguistic Perspective*, ed. Bernd Kortmann, 335-363. Berlin: Mouton de Gruyter.

Rohdenburg, Günter. 2004b. Grammatische Parallelen zwischen niederdeutschen Mundarten und Nichtstandardvarietäten im Englischen aus typologischer Sicht. *Niederdeutsches Jahrbuch* 127:85-122.

Romijn, Kirsten. 1996. *Hoe doen we het? Verwijzen naar linguïstische en cognitieve representaties met het voornaamwoord "het", Academisch proefschrift P.J. Meertens-Instituut*. Amsterdam: Publicaties van het P.J. Meertens-Instituut.

Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. In *Syntax. Ein internationales Handbuch zeitgenössischer Forschung/ An International Handbook of Contemporary Research*, eds. JoachimJacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann, 646-686. Berlin: de Gruyter.

Scholtz, J. du P. 1963. Die Ondergang van die Nominale Tweeklassesisteem in Afrikaans. In *Taalhistoriese Opstelle*, ed. J. du P. Scholtz. Pretoria: van Schaik.

Siemund, Peter. 2002a. Mass versus count: Pronominal gender in regional varieties of Germanic languages. *Zeitschrift für Sprachtypologie und Universalienforschung (STUF)* 55:213-233.

Siemund, Peter. 2002b. Animate Pronouns for Inanimate Objects. Pronominal Gender in English Regional Varieties. In *Anglistentag 2001 Vienna*, ed. Dieter Kastovsky, 19-34. Trier: Wissenschaftlicher Verlag.

Siemund, Peter. 2008. *Pronominal Gender in English - A study of English varieties from a cross-linguistic perspective*. London: Routledge.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Sijs, Nicoline van der. 2003. De herkomst van de haar-ziekte. *Onze Taal* 10:268-269.

Sijs, Nicoline van der. 2004. Het woordgeslacht van de zelfstandige naamwoorden en wat daarmee samenhangt. In *Het ontstaan van het ABN*, ed. ibid. Amsterdam: Athenaeum.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, ed. Richard Dixon, 112-171. Canberra: Australian Institute of Aboriginal Studies.

Strandskogen, Åse-Berit, and Strandskogen, Rolf. 1995. *Norwegian: An Essential Grammar*. London: Routledge.

Strauss, Sabine. 2007. *The Gender of Pronouns Co-Occurring With Neuter Hybrid or Epicene Antecedents in German*: unpublished MA thesis, VU University Amsterdam.

Svartengren, T. Hilding 1927. The feminine gender for inanimate things in Anglo-American. *American Speech* 3: 83-113.

Tasmowski, Liliane, and Verluyten, S. Paul 1985. Control mechanisms of anaphora. *Journal of Semantics* 4:341-370.

Tasmowski-De Ryck, Liliane, and Verluyten, S. Paul 1982. Linguistic control of pronouns. *Journal of Semantics* 1:323-346.

Kate, Lambert H. Ten, Noordegraaf, Jan, and Van der Wal, Maria J. 2001 [1723]. *Aenleiding tot de kennisse van het verhevene deel der Nederduitsche sprake*. Alphen aan den Rijn: Canaletto/Repro Holland.

Timberlake, Alan. 1975. Hierarchies in the Genitive of Negation. *The Slavic and East European Journal* 19:123-138.

Verhoeven, P.R.F. 1990. Voornaamwoordelijke aanduiding in het hedendaagse Nederlands. *De nieuwe taalgids* 83:494-513.

Viejo Fernández, Xulio. 2001. Metafonía y contrametafonía en quirosán: pragmática, determín léxicu y morfosintaxis. *Revista de Filología Asturiana* I.

Vogel, Petra M. 2000. Nominal abstracts and gender in Modern German: A "quantitative" approach towards the function of gender. In *Gender in Grammar and Cognition*, eds. Barbara Unterbeck and Matti Rissanen, 461-494. Berlin: Mouton de Gruyter.

Vogelaer, Gunther de. 2006. Pronominaal genus bij 'Zuid-Nederlandse' taalverwervers: van grammaticaal naar semantisch systeem. In *Nederlands tussen Duits en Engels*, eds. Matthias Hüning, Ulrike Vogl, Ton van der Wouden and Arie Verhagen. Leiden: Stichting Neerlandistiek Leiden.

Vogelaer, Gunther de. to appear a. (Not) acquiring grammatical gender in two varieties of Dutch. In *Recent advances in Cognitive Sociolinguistics*, eds. Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman. Berlin: Mouton de Gruyter.

Vogelaer, Gunther de. to appear b. Frequency effects in gender change: data from East and West Flemish dialects. In *Papers of the Linguistic Society of Belgium. 3*, ed. Lobke et al. Aelbrecht. Brussels: Linguistic Society of Belgium.

Vogelaer, Gunther de, and Vos, Lien de. submitted. *Dutch gender and the locus of morphological regularisation*. Ms. UGent/Université de Liège.

Vries, Jelle de. 2001. *Onze Nederlandse spreektaal*. Den Haag: Sdu Uitgevers.

Vries, M. de, Winkel, L.A. te. 1866. *Woordenlijst voor de spelling der Nederlandsche taal*. 's Gravenhage: Nijhoff.

Wagner, Susanne. 2003. *Gender in English pronouns: myth and reality*. PhD thesis University of Freiburg (Breisgau).

Wahrig-Burfeind, Renate. 1989. *Nominales und pronominales Genus im südlichen Nordseegebiet. Eine areallinguistische Untersuchung*. München: tuduv.

Wal, Marijke J. Van der. 1988. Latijn leren en de moedertaal observeren: grammatikale opmerkingen over de volkstaal in het Exercitium puerorum 1485. *Gramma* 12:243-257.

Wechsler, Steven, and Zlatić, Larisa. 2003. *The Many Faces of Agreement*. Stanford: CSLI Publications.

Wiesemann, Ursula. 1986. *Pronominal systems*. Tübingen: Narr.

Williams, Edwin. 1981. On the Notions 'Lexically Related' and 'Head of a Word'. *Linguistic Inquiry* 12:245-274.

Williamson, Kay and Roger Blench. 2000. Niger-Congo. In *African Languages: An Introduction*, Bernd Heine and Derek Nurse (eds.), Cambridge: Cambridge University Press. 11–42.

Williamson, Kay. 1965. *A grammar of the Kolokuma dialect of Ijo*. Cambridge: Cambridge. University Press.

Wurzel, Wolfgang U. 1986. Die wiederholte Klassifikation von Substantiven. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation* 39:76-96.

Zaliznjak, Andrej A. 1973. O ponimanii termina 'padež' v lingvističeskix opisanijax. In: Andrej A. Zaliznjak (ed.) *Problemy grammatičeskogo modelirovanija*. Moscow: Nauka, 53-87. [Reprinted in: Andrej A. Zaliznjak. 2002. *Russkoe imennoe slovoizmenenie: s priloženiem izbrannyx rabot po sovremennomu russkomu jazyku i obščemu jazykoznaiju*. Moscow: Jazyki slavjanskoj kultury. 613-47.]

Zwicky, Arnold M. 1969. Phonological constraints in syntactic descriptions. *Papers in Linguistics* 1:411-453.

Zwicky, Arnold M. 1992. Some Choices in the Theory of Morphology. In *Formal Grammar Theory and Implementation*, ed. Robert Levine, 327-370. Oxford: Oxford University Press.

Zwicky, Arnold M., and Pullum, Geoffrey K. 1986. The Principle of Phonology-Free Syntax: introductory remarks. *Working Papers in Linguistics, Ohio State University* 32:63-91.

Grammars and Dictionaries (electronic editions)

Van Dale Groot Woordenboek van de Nederlandse Taal
        http://vandale.nl
Algemene Nederlandse Spraakkunst
        http://www.let.ru.nl/ans/e-ans/
Woordenlijst Nederlandse Taal ("Het Groene Boekje")
        http://woordenlijst.org

De heruitvinding
van het pronominale genus

Dit boek beschrijft een empirisch onderzoek naar het pronominale genussysteem van het hedendaags gesproken Nederlands. Het Nederlands heeft een interessante tegenstrijdigheid: de lidwoorden, de bijvoeglijke naamwoorden en de betrekkelijke voornaamwoorden onderscheiden twee geslachten, terwijl de voornaamwoorden een vierledig onderscheid maken, namelijk tussen *hij* (masculien), *zij* (feminien), *het* (neutraal) en *die* (commuun). Dit staat in tegenspraak met onze verwachtingen omtrent agreementsystemen: het naamwoord en alle daarmee congruerende elementen hebben normaalgesproken dezelfde genuswaarden. In het Nederlands gaat deze mismatch hand in hand met een verbazingwekkende variatie in het gebruik van voornaamwoorden. *De*-woorden vindt men met masculiene, feminiene en neutrale voornaamwoorden, terwijl *het*-woorden regelmatig door *hij*, *zij*, *deze* of *die* worden gevolgd. Ook dit staat in tegenspraak met de verwachtingen. Immers, agreement wordt gedefinieerd als een overeenkomst van grammaticale eigenschappen.

De situatie is geworteld in de geschiedenis van het Nederlands. Rond de tijd van het Middelnederlands ging het onderscheid tussen het masculinum en het femininum in de lidwoorden en de bijvoeglijke naamwoorden verloren. Na verloop van tijd verdween ook het 'genusgevoel', de kennis welke niet-onzijdige woorden bij welk geslacht horen. Van officiële zijde werden woordenlijsten samengesteld die het behoud van de oorspronkelijke geslachten veilig moesten stellen. Voor het correcte gebruik van voornaamwoorden werden sprekers geacht het woordenboek te raadplegen. Dit veroorzaakte een groeiende kloof tussen spreektaal en schrijftaal. Ook tegenwoordig passen taalgebruikers verschillende strategieën toe bij het gebruik van voornaamwoorden in spraak en schrift. Deze studie had tot doel om de spreektalige strategieën in kaart te brengen.

De theoretische benadering ging uit van de aanname dat voornaamwoorden agreement met hun antecedent vertonen, op dezelfde manier als lidwoorden congrueren met hun zelfstandig naamwoord. Dus, voornaamwoorden met het 'verkeerde' genus werden niet als buiten het systeem vallend beschouwd, maar gezien als gevallen van overeenkomst met andere eigenschappen van het naamwoord. Naar aanleiding van onze kennis over morfosyntaxis en agreement werd er verwacht dat deze eigenschappen semantisch van aard zijn.

Corpusonderzoek op basis van 500.000 woorden spontane spraak wijst uit dat Nederlandse sprekers hun voornaamwoorden inderdaad kiezen op basis van semantiek. Masculiene pronomina worden gebruikt in verwijzing naar mannelijke personen, naar dieren (vaak zelfs naar vrouwelijke dieren) en naar telbare, begrensde, specifieke voorwerpen en abstracta. Neutrale (onzijdige) voornaamwoorden daarentegen worden voor stofnamen gebruikt. Feminiene

231

pronomina kennen het sterkst beperkte gebruik. Zij komen alleen voor in verwijzing naar vrouwelijke personen en dieren.

Alle voornaamwoorden die niet overeenkwamen met het geslacht van hun antecedent bleken in deze semantische patronen te passen. Gecombineerd ontstaat er een homogeen beeld van het pronomengebruik in het hedendaags gesproken Nederlands. Voornaamwoorden zijn verdeeld volgens de *individueringsgraad* van de referent. Het meest geïndividueerd zijn personen, gevolgd door dieren en telbare objecten. Dit is het domein van het masculinum (met uitzondering van vrouwelijke personen en dieren, waarvoor het femininum wordt gebruikt). Als het laagst geïndividueerd gelden stoffen en niet-telbare abstracta die met het onzijdige genus worden geassocieerd. Ertussenin ligt de klasse van de specifieke stoffen. Dit zijn gevallen als *mijn thee* of *deze wijn*. Deze groep combineert eigenschappen van zaaknamen en stofnamen. Hier kunnen masculiene en neutrale pronomina voorkomen.

De individuering van een referent hangt af van de *construal* van de situatie. Een vis kan als individueel (huis)dier worden gezien, maar ook als onspecifiek onderdeel van een school of zelfs als ingrediënt van een maaltijd. De gekozen voornaamwoorden verschillen: in het eerste en tweede geval zal een spreker mannelijke pronomina kiezen, in het derde onzijdige. Dit verklaart de variatie tussen de voornaamwoorden voor hetzelfde zelfstandige naamwoord. Variatie treedt op tussen sprekers, maar ook in de spraak van een individueel persoon. Zelfs in de schrijftaal, die normaliter syntactische congruentie voorschrijft, zijn er voorbeelden te vinden.

Op deze wijze wordt aangetoond dat achter het schijnbaar chaotische voornaamwoordgebruik in het Nederlands een regelmatig en logisch systeem schuilgaat. Dit systeem heeft parallellen in andere talen, zowel in het Germaans als ook erbuiten.

Naast het semantische systeem gebruiken sprekers ook nog steeds het traditionele systeem van syntactische congruentie. Dit systeem combineert *het*-woorden met de voornaamwoorden *het* of *dat*, en *de*-woorden met de aanwijzende voornaamwoorden *deze* of *die*. De keuze tussen het ene en het andere systeem wordt door een aantal factoren beïnvloed. Een statistische analyse van de corpusdata toont aan dat de volgende factoren een rol spelen in het keuzeproces. Semantisch agreement is waarschijnlijker

- voor persoonlijke dan voor betrekkelijke voornaamwoorden
- voor volle vormen (*hij, zij, het*) dan voor gereduceerde vormen (*ie, ze, 't*)
- voor pronomina in de nominativus (*hij, zij*) dan voor pronomina in de dativus/accusativus (*hem, haar*)
- voor referenten aan het uiteinde van de Individueringshiërarchie (personen, stoffen) dan voor referenten in het midden (objecten)

De waarschijnlijkheid dat een spreker voor semantisch agreement kiest, neemt toe, naarmate voornaamwoord en antecedent verder van elkaar verwijderd zijn. Als een keer een semantisch pronomen is gekozen, keert men zelden terug naar syntactisch agreement. Tot slot blijkt de leeftijd van de spreker van invloed te zijn. Sprekers van onder de 20 gebruiken dubbel zo vaak semantisch gemotiveerde voornaamwoorden dan sprekers van boven de 60. Dit suggereert dat het semantische systeem zich uitbreidt ten koste van het traditionele, syntactische systeem.

Met behulp van de Individueringshiërarchie en de Agreement Hierarchy kan de variatie in het voornaamwoordgebruik worden verklaard. De oorzaak ligt in de meeste gevallen in verschillende *construals* van de situatie of in de competitie tussen het oude en het nieuwe systeem.

Ook in andere Germaanse talen blijken vergelijkbare systemen te bestaan. Een overzicht van de standaardtalen geeft een verrassend beeld: meer dan de helft heeft meer pronominale dan attributieve geslachten. Dit geldt voor het Engels, het Afrikaans, het Deens, het Zweeds, variëteiten van het Noors en het Fries. In alle gevallen blijken de pronominale geslachten op semantische basis gebruikt te worden, zelfs als de rest van het genussysteem niet primair semantisch is. Dit suggereert dat pronomina niet in staat zijn om een syntactisch onderscheid op eigen kracht overeind te houden. Dit verklaart waarom het (Noord-)Nederlands geen masculiene en feminiene zelfstandige naamwoorden meer kent. Toen de attributieve markering verdween en alleen de voornaamwoorden overbleven, konden deze het traditionele systeem niet voldoende houvast geven.

Deze aanname leidt tot de hypothese dat bepaalde soorten agreement betere steun aan een genussysteem bieden dan andere. Om na te gaan of dit crosslinguïstisch juist is, werd evidentie van talen verzameld die genus alleen op het persoonlijke voornaamwoord markeren. 20 relevante talen werden gevonden, uit alle delen van de wereld. In vrijwel alle gevallen bleek het systeem inderdaad strikt semantisch georganiseerd. Het is aannemelijk dat meer complexe systemen, die niet alleen semantische, maar ook fonologische en morfologische genusregels kennen, meer steun door agreement nodig hebben, bij voorkeur binnen het locale domein van de NP.

In dit licht blijken de buitengewone feiten van het voornaamwoordgebruik in het Nederlands goed aan te sluiten bij diverse typologisch vertrouwde verschijnselen. Dit bevestigt de natuurlijkheid van de ontwikkeling van een op syntaxis gebaseerd naar een semantisch pronominaal genussysteem.

Voor de onbevooroordeelde beschouwer toont zich in de pronominale 'fouten' in spontane spraak een ingenieus en logisch systeem dat op natuurlijke wijze het pronominale probleem van het Nederlands oplost.