BILINGUAL ACQUISITION
OF GENDER
AND GENDER AGREEMENT

by

Magdalena Fiałkowska

Submitted for the Degree of Doctor of Philosophy

Surrey Morphology Group
Department of English
Faculty of Art and Human Science
University of Surrey

January 2011
This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully indentified as such within the work and are attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

Magdalena Fiałkowska
University of Surrey
January 2011
Abstract

The scope of this thesis is grammatical gender. It investigates the speech of three children raised bilingually in two Indo-European languages, Polish and English, and presents a plethora of aspects that need to be taken into account when studying the acquisition of gender by children. Acquisition of two languages simultaneously is a complex process and has been studied by scholars from many different perspectives. This study investigates only one element of the children's linguistic system, but does it with care and in detail. First, the nature and function of gender in Polish and English are discussed; second, a review of studies dealing with acquisition of grammatical gender in different languages is provided, and accompanied by short debate on language development theories. Next, the reader is presented with details on the methodology implemented for this project, i.e. recruitment procedures, data collection process, coding techniques and more. This is followed by the data analysis, which is focused on two areas: Polish noun phrases and Polish-English mixed noun phrases. Finally, the thesis is concluded with a list of findings and ideas for future research. This thesis contributes to the following areas of research: linguistics, child language, bilingualism and language contact. It is the author's own work, and all data come from the author's own database, collected from the children and their families in the United Kingdom between late 2006 and early 2009.
# Table of contents

ABSTRACT ................................................................................................................................................... IV

TABLE OF CONTENTS .............................................................................................................................. V

LIST OF FIGURES ..................................................................................................................................... VII

LIST OF TABLES ......................................................................................................................................... IX

LIST OF ABBREVIATIONS ....................................................................................................................... X

ACKNOWLEDGEMENTS ........................................................................................................................ XI

INTRODUCTION ......................................................................................................................................... 12

CHAPTER ONE. WHAT IS GENDER? ................................................................................................. 14

1.0 THE NATURE OF GENDER ........................................................................................................... 14

1.0.1 Functions of gender .................................................................................................................. 14

1.0.2 Agreement .................................................................................................................................. 15

1.0.3 Gender assignment .................................................................................................................... 19

1.0.3.1 Semantic criteria ................................................................................................................... 19

1.0.3.2 Morphological criteria ........................................................................................................ 21

1.0.3.3 Phonological criteria ............................................................................................................ 22

1.1 GENDER IN POLISH ................................................................................................................... 23

1.1.1 The nature of gender in Polish .................................................................................................. 23

1.1.2 Gender assignment ................................................................................................................... 25

1.1.2.1 Semantic criteria ................................................................................................................... 25

1.1.2.2 Formal criteria ..................................................................................................................... 27

1.1.3 How many genders in Polish? .................................................................................................. 30

1.1.4 Agreement .................................................................................................................................. 37

1.2 GENDER IN ENGLISH .................................................................................................................. 44

1.2.1 Gender assignment ................................................................................................................... 45

1.2.2 Gender agreement .................................................................................................................... 47

1.3 CONCLUSIONS ............................................................................................................................ 48

CHAPTER TWO. ACQUISITION OF GENDER ................................................................................. 49

2.0 INTRODUCTION ........................................................................................................................... 49

2.1 LANGUAGE DEVELOPMENT ...................................................................................................... 49

2.2 BILINGUAL LANGUAGE ACQUISITION ..................................................................................... 55

2.3 ACQUISITION OF GRAMMATICAL GENDER ........................................................................... 59

2.4 CONCLUSIONS ............................................................................................................................ 72
CHAPTER THREE. METHODOLOGY .............................................................................................................. 74

3.0 INTRODUCTION ................................................................................................................................. 74
3.1 RECRUITMENT ................................................................................................................................. 74
3.2 THE INFORMANTS ........................................................................................................................... 78
   3.2.1 Family 1: Patrick ....................................................................................................................... 79
   3.2.2 Family 2: Jerzy ......................................................................................................................... 82
   3.2.3 Family 3: Hania ....................................................................................................................... 87
3.3 MLU ....................................................................................................................................................... 91
3.4 DATA COLLECTION ........................................................................................................................... 95
3.5 CORPORA .......................................................................................................................................... 99
   3.5.1 CHAT ........................................................................................................................................ 100
   3.5.2 CLAN ......................................................................................................................................... 104
   3.5.3 MOR ......................................................................................................................................... 107
3.6 CONCLUSIONS .................................................................................................................................. 112

CHAPTER FOUR. DATA ANALYSIS ........................................................................................................ 114

4.0 INTRODUCTION .................................................................................................................................. 114
4.1 LANGUAGE PROFILES ..................................................................................................................... 114
4.2 DATA .................................................................................................................................................. 120
   4.2.1 Polish data .............................................................................................................................. 120
       Analysis 1 Diachronic error shift ................................................................................................. 121
   4.2.2 Mixed data ............................................................................................................................. 158
       Analysis 1 Mixed data .................................................................................................................. 163
       Analysis 2 Various tendencies ..................................................................................................... 167
       Analysis 3 Which clues? ............................................................................................................... 176
       Analysis 4 Inflection ..................................................................................................................... 184
4.3 CONCLUSIONS .................................................................................................................................. 186

CHAPTER FIVE. RESULTS AND IMPLICATIONS ..................................................................................... 187

5.0 INTRODUCTION ............................................................................................................................... 187
5.1 FINDINGS AND IMPLICATIONS ...................................................................................................... 188
5.2 FUTURE RESEARCH ........................................................................................................................ 196

REFERENCES ............................................................................................................................................. 198
List of figures

Figure 1. Adam, Sarah and Eve (Brown, 1973) ......................................................................................... 75
Figure 2. MOR library .................................................................................................................................. 108
Figure 3. Children's error frequencies ........................................................................................................ 125
Figure 4. Feminine replaced by masculine or neuter (Hania) .................................................................. 127
Figure 5. Feminine replaced by masculine or neuter (Patrick) ................................................................. 128
Figure 6. Feminine replaced by masculine or neuter (Jerzy) ..................................................................... 128
Figure 7. Masculine replaced by feminine or neuter (Hania) .................................................................... 130
Figure 8. Masculine replaced by feminine or neuter (Patrick) ................................................................. 130
Figure 9. Masculine replaced by feminine or neuter (Jerzy) ..................................................................... 131
Figure 10. Feminine as the replacing gender .............................................................................................. 133
Figure 11. Masculine as the replacing gender ............................................................................................ 133
Figure 12. Neuter as the replacing gender .................................................................................................. 134
Figure 13. Masculine animate replaced by masculine inanimate ................................................................ 136
Figure 14. Masculine inanimate replaced by masculine animate ............................................................. 137
Figure 15. Masculine animate-inanimate noun frequency (2;4-4;3) .......................................................... 138
Figure 16. Mapping of the overall animate-inanimate noun frequencies to animate-inanimate gender errors .......................................................................................................................................................... 141
Figure 17. Overall correct-incorrect agreement in masculine and feminine NPs at 3;1-3;4 (1) .................. 143
Figure 18. Overall correct-incorrect agreement in masculine and feminine NPs at 3;1-3;4 (2) ............... 143
Figure 19. Masculine correct-incorrect agreement in detail (3;1-3;4) .................................................... 144
Figure 20. Feminine correct-incorrect agreement in detail (3;1-3;4) ....................................................... 144
Figure 21. Overall correct-incorrect masculine animate-inanimate agreement at 3;1-3;4 .................... 145
Figure 22. Correct masculine animate-inanimate agreement in detail (3;1-3;4) .................................... 145
Figure 23. Incorrect masculine animate-inanimate agreement in detail (3;1-3;4) ................................. 146
Figure 24. Animate-inanimate nouns at 3;6;3;8;3;10;4;1 ..................................................................... 147
Figure 25. Overall case frequencies for masculine nouns ........................................................................ 147
Figure 26. Feminine replaced by masculine in the accusative and non-accusative cases ..................... 149
Figure 27. Masculine personal/non-masculine personal mutual replacements ........................................ 152
Figure 28. Errors across number ............................................................................................................... 155
Figure 29. The frequency of mixed NPs ..................................................................................................... 164
Figure 30. The frequency of mixed utterances ........................................................................................... 165
Figure 31. Hania's mixing rates .................................................................................................................. 166
Figure 32. Patrick's mixing rates ............................................................................................................... 166
Figure 33. Jerzy's mixing rates .................................................................................................................... 166
Figure 34. Genders in mixed NPs ................................................................................................................ 167
List of tables

TABLE 1. ELEMENTS TAKING AGREEMENT IN POLISH .......................................................... 31
TABLE 2. ADJECTIVAL PARADIGMS IN POLISH – SINGULAR (SCHENKER, 1955) .................. 34
TABLE 3. ADJECTIVAL PARADIGMS IN POLISH – PLURAL (SCHENKER, 1955) ..................... 34
TABLE 4. NOUN PARADIGMS IN POLISH ............................................................................ 36
TABLE 5. PREDICATIVE AGREEMENT .................................................................................. 37
TABLE 6. PREDICATIVE AND ATTRIBUTIVE AGREEMENT ..................................................... 37
TABLE 7. ATTRIBUTIVE AGREEMENT IN POLISH ................................................................ 38
TABLE 8. PREDICATIVE AGREEMENT IN POLISH .................................................................. 39
TABLE 9. RESEARCH IN THE ACQUISITION OF GENDER ........................................................ 69
TABLE 10. PATRICK’S PLACES OF RESIDENCE ..................................................................... 80
TABLE 11. PATRICK’S LINGUISTIC ENVIRONMENT AT RECORDING SESSIONS ................. 82
TABLE 12. JERZY’S PLACES OF RESIDENCE .......................................................................... 84
TABLE 13. JERZY’S LINGUISTIC ENVIRONMENT AT RECORDING SESSIONS ......................... 85
TABLE 14. HANNA’S PLACES OF RESIDENCE ......................................................................... 89
TABLE 15. HANNA’S LINGUISTIC ENVIRONMENT AT RECORDING SESSIONS ......................... 89
TABLE 16. SPEAKER CODES .................................................................................................. 104
TABLE 17. CLAN CODING COMPARED WITH SMOCZYNSKA’S CODING .................................. 111
TABLE 18. MLU AT THE AGE OF 29 MONTHS ......................................................................... 117
TABLE 19. MLUS AT THE AGE OF 29 MONTHS ....................................................................... 118
TABLE 20. TTR AT THE AGE OF 29 MONTHS ......................................................................... 119
TABLE 21. MLT AT THE AGE OF 29 MONTHS ......................................................................... 120
TABLE 22. GENERAL ERROR CLASSIFICATION ..................................................................... 122
TABLE 23. DETAILED ERROR CLASSIFICATION ................................................................... 123
TABLE 24. ERROR TYPES ACCORDING TO FREQUENCY ....................................................... 124
TABLE 25. GENDER ACCURACY ............................................................................................ 180
TABLE 26. INFLECTIONAL SUFFIXES ON ENGLISH INSERTIONS (SG) ................................. 185
TABLE 27. INFLECTIONAL SUFFIXES ON ENGLISH INSERTIONS (PL) ................................. 185
List of abbreviations

1 first person
2 second person
3 third person
ACC accusative
ADJ adjective
DAT dative
F feminine
FUT future
GEN genitive
IMP imperative
INF infinitive
INS instrumental
IPFV imperfective
LOC locative
M masculine
M.ANIM masculine animate
M.INAN masculine inanimate
M.PERS masculine personal
N neuter
N_ non
N_M.PERS non-masculine personal
NOM nominative
PFV perfective
PL plural
PROG progressive
PRS present
PST past
Q question
REFL reflexive
SG singular

The Leipzig Glossing Rules are the main source of reference for all glosses used in the thesis. They can be found at http://www.eva.mpg.de/lingua/resources/glossing-rules.php.
Acknowledgements

This project would not have been possible without the time and effort from the participating families. Thank you for being so welcoming, friendly and for giving me faith and optimism. To all the children – Amelia, Harry, Jaś, Jennifer, Malcolm, Mikolaj, Maks and Ola – thank you for your openness and acceptance. Your input in this project is priceless.

This thesis would not have been written without the help of my supervisors, Greville Corbett and Dunstan Brown, who encouraged and challenged me throughout my academic journey. They guided me never accepting less than my best efforts. I wish to thank Magdalena Smoczyńska for her generosity in offering time, expertise and her own research tool, which enabled the coding and analysis of my data. You have my honest thanks. To Arjen Versloot for his genuine interest, academic spirit, and knowledge. Also, to both Elena Lieven and Lydia White for their constructive feedback.

To The Surrey Morphology Group for their inspiring work and support: Matthew Baerman, Marina Chumakina, Sasha Krasovitsky, Ali Long, Claire Turner – thank you for sharing your thoughts with me throughout those years. You have shown me the standards of the best quality academic work. To Anna Kibort for all the thought-provoking and brainstorming conversations, which showed me new solutions and perspectives. You have my very special thanks. Also, to Napoleon Katsos for providing helpful comments, proofreading parts of this thesis and sharing his honest opinion with me. Thank you also to the other academic and administrative staff in the department for understanding, encouragement and help.

To the Polish Ministry of Education and the University of Surrey, thank you for the financial support that made my fieldwork possible and enabled me to spend four productive years in the UK working on this project.

Last but not the least, I wish to thank all my family and friends in both Poland and England, who never ceased believing in me. Especially, I want to thank my parents, Danuta and Tadeusz Fialkowski for their wonderful love. To my siblings, Joanna and Artur, who gave me strength to fight for my ambitions. To my landlords, Felicity and Glyn Radcliffe-Brine for being like a true family to me by showing interest and never failing to encourage me in times of doubt. You are wonderful. To St Joseph’s Worship Band for your spiritual support. To Ula Polanska, Chris Shedd and Mike Longridge, who walked many extra miles to accompany me in this journey to give their love and true friendship. To my dear friends, Judy and Chris Collins and family, Nicki and Alan Lion and family, Lillian Simms and family, Aseel Gilbert, Tim Hunting, Louis Nyuyse, Rich Milner and many more, whose contribution, large or small, is greatly appreciated.

You all have my honest and endless thanks.
Introduction

This is a study based on raw data collected from three children: Hania, Jerzy and Patrick (names of all the participants have been altered to ensure their full anonymity and protection as research subjects). Studying children between ages 2;5 and 4;3 gives an enormous range of options, which may take a lifetime if one wants to exhaust them in detail. The aim of this study is to provide fresh and pioneering bilingual data, which will contribute to research conducted in child language and bilingualism. In contrast to English, which is one of the languages studied here, Polish has not been investigated enough so far. Taken into consideration the intriguing structure of Polish and the intricacies of the system, the amount of work carried out in Poland and abroad is still insufficient. In addition, the rapidly changing socio-economic reality of Poland has increased the awareness of the value of bilingual upbringing of children and adults. Parents believe in assisting and investing in their children's learning of another language from an early age. These are standard attitudes in modern Poland. For Polish or Polish-English families living abroad, supervising the child's learning of both languages to a native-like level is also very often a matter of great importance. The view shared by many Polish-English families is that under no circumstances should the Polish language be abandoned or treated as less valuable or important than English. In this context, parents appreciate a better understanding of how to assist their children in learning Polish when they are also acquiring English as the majority language. This project is not a socio-linguistic study; neither does it deal with issues of biculturalism. It is a strictly linguistics-oriented investigation, which focuses on the three children's learning of grammatical gender in Polish, which co-exists in their linguistic repertoire with English. Both languages are the children's first languages and neither is prioritised over the other. This work is important for a few reasons. Firstly, studying the acquisition of grammatical gender in two languages so different with respect to gender allows us access to mechanisms which children employ to distinguish the two systems. Secondly, studying bilinguals contributes important data and insight into language learning which is not available from the studies on monolinguals. Finally, uncovering error patterns in learning gender by natural bilinguals may also shed light on the processing and understanding of gender in L2 learning.
The structure of the thesis is as follows. Chapter One introduces grammatical gender and discusses its nature and functions in Polish and English. This chapter also defines gender assignment and gender agreement. Chapter Two follows with a review of studies dealing with the acquisition of grammatical gender in different languages. Firstly, a short discussion on language development theories is provided, and afterwards, the bilingual acquisition of gender is discussed. In Chapter Three, the reader will find all necessary details pertaining to the methodology implemented in this project. The recruitment procedures, the data collection process and the coding techniques are described. Moreover, all the participants are introduced there: their linguistic environment and the communication patterns at home are portrayed in all necessary detail. Chapter Four is focused solely on the data analysis. It begins with language profiles of the three participating children, and proceeds to two main sections of this chapter: the Polish error data, and the mixed Polish-English NPs. The Polish error data discusses diachronic error shifts in the singular and in the plural, as well as across number. The mixed Polish-English data deals with aspects such as animacy or masculine as the default. Finally, all results are presented in Chapter Five, which ends with a list of ideas for future research.
Chapter One. What is gender?

1.0 The nature of gender

1.0.1 Functions of gender

Although gender as an area of study gives many opportunities for linguistic research, it is not a universal category. If we compare some functions of gender, such as showing the attitude of the speaker, i.e. marking status, showing respect or lack of it and displaying affection (Corbett, 1991: 320-323), we must admit that those functions available via gender distinctions are not vital for the proper functioning of any language. The nature of gender has been described as “la distinction des genres est un luxe linguistique, sans relation avec la logique” (Bally, 1935: 45 after Fyodor, 1959: 195), which dooms gender to be an unessential category, with no useful purpose observable by other means. This gloomy approach can be contrasted with a more optimistic view of the functions of natural and grammatical gender presented based on a wide range of languages.

Gender distinctions have a number of central functions that are based on the distinctions male : female or animate : inanimate (in the indication of the sex or animacy of a human or animal). This is certainly true for Polish. In his discussion on the functions of gender, when referring to grammatical gender in Polish, Trudgill uses the term “arbitrary”, while Kilarski views gender assignments as “tendencies”, supporting his argument with the role of gender in the organisation of the lexicon (Kilarski and Trudgill, 2000: 193ff). It can be observed how Trudgill’s critical view on the functions of gender moderates from claiming that gender as a category can be explained but it has no function to admitting that if the grammatical gender has any functions, they are “relatively minor and peripheral”. Importantly, Trudgill also points out that “the semantic basis of gender at least in modern Indo-European languages is deeply opaque and probably largely non-existent”, which may be why he considers Polish gender assignment rules highly arbitrary. In conclusion, Kilarski

1 “The distinction of genders is a linguistic luxury, without relation with the logic” [translation mine]
and Trudgill agree that if we consider all these minor functions, such as indicating speaker's attitude, reference tracking, personification in literature or free word order (e.g., in Persian), "we get a rather opaque picture of the functions of natural and grammatical gender".

A question remains whether natural and grammatical gender distinctions are merely "a luxury" to have in a language. Some may say it is more of a burden than a luxury during the acquisition period. Whether it really is a burden for young language learners will be examined based on the data collected for the purpose of this thesis.

1.0.2 Agreement

Defining agreement is a daunting task, and it is even more challenging to provide a general, considerably full, yet brief description of this phenomenon, especially if one considers such thorough studies of agreement as Lehmann (1982), Barlow (1988) or Corbett (2006). A selection of definitions is presented below:

The term agreement commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. For example, adjectives may take some formal indication of the number and gender of the noun they modify (Steele, 1978: 610).

A grammatical constituent A will be said to agree with a grammatical constituent B in properties C in language L if C is a set of meaning-related properties of A and there is a covariance relationship of between C and some phonological properties of a constituent Bi across some subset of the sentences of language L, where constituent Bi is adjacent to constituent B and the only meaning related non-categorial properties of constituent Bi are the properties C (Moravcsik, 1978: 1).

Moravcsik’s “agreeing constituents” are referred to as “elements” in Steele’s definition, and they may be for instance subject NP and a verb in English (Little Mike is my nephew) or a verb agreeing with a subject NP plus a modifying adjective in Polish (Rower Jasia byl zielony ‘Johnny’s bike was green’). In the English example, the verb (constituent/element) agrees in person and number, “because there is a relationship of covariance between the number and person specifications of the subject noun phrase and between the phonological shape of the verbal suffix” (Moravcsik, 1978: 1). The same may be said about the agreement between the Polish NP Rower Jasta, the verb byl and the modifying adjective zielony, where there is a relationship of covariance between the number and person specifications of the

Fodor, 1959:206: "When for instance in the Russian folk song "Pesn’a o r’abine" (Song about the Rowan) r’abina fem. 'rowan' is yearning for the distant dub masc. 'oak', the feminine gender of r’abina determines the content as well as mood of the poem, and constitutes a stylistic device, of which genderless language would be incapable, the same thought and feeling having to be rendered by different means or in roundabout ways".
subject NP and the phonological shape of the adjectival and verbal suffixes. An additional “property” in the Polish example is the agreement in gender, which is evident through the choice of the masculine suffix -y in the adjective zielony, and the suffix -l in the verb byl that is characteristic for the masculine gender for Polish verbs in the past tense.

Agreement in Polish is marked on the following parts of speech: adjectives (attributive and predicative), verbs, demonstratives, pronouns (personal, possessive and relative), numerals, participles, determiners, adverbs and complementizers. The above-quoted definitions show that it is based on the interrelation between different elements both within a sentence and beyond, and this interrelation is the essence of gender. There are two reasons for this claim: firstly, agreement is the sole evidence for demonstrating the existence of gender, and secondly, agreement is the basis for determining the number of genders in a given language. The evidence is based on agreement markers “attached to other sentence elements, whose form is determined by the gender of the head noun of the controller” (Corbett, 1991: 147). In other words, gender classes (nouns) can only be distinguished by the agreement they take. An extensive analysis of agreement in different world languages is provided in Corbett (1991). This meticulous study shows a variety of domains in which agreement takes place, as well as how the agreement class approach leads us to differentiate controller genders (into which nouns are divided), from target genders (marked on adjectives, verbs etc) (Corbett, 1991: 151). These two elements involved in agreement have been defined and framed in the context of agreement:

We call the element which determines the agreement the controller. The element whose form is determined by agreement is the target. The syntactic environment in which agreement occurs is the domain of agreement. And when we indicate in what respect there is agreement, we are referring to agreement features (Corbett, 2001: 2).

Corbett’s (1991: 147) definition is based on the idea found in Zaliznjak (1964):

An agreement class is a set of nouns such that any two members of that set have the property that whenever

(i) they stand in the same morphosyntactic form  
and  
(ii) they occur in the same domain  
and  
(iii) they have the same lexical item as agreement target  
then  
their targets have the same morphological realisation

Such a definition requires at least a brief comment. If two nouns stand in the same morphosyntactic form, their syntactic features, e.g., number and case, are the same. In the
simplest terms, it means that the nouns involved should be in the same number and the same case. However, it does not mean that the same morphosyntactic form implies identical morphological form, i.e., the nouns involved may for instance take different endings. Next, if two nouns occur in the same domain, it means that “the environment” is identical in each case, i.e. it is for instance subject-verb agreement. Finally, two nouns have the same agreeing targets if both those targets show for example agreement in gender or/and distinguish the same number of genders (Corbett, 1991: 148).

Various linguists, e.g., Moravcsik (1978: 334), Lehmann (1982), Givón (1976: 151), Lapointe (1985) and Corbett (1991: 112), support the non-local treatment of agreement including anaphoric pronouns within its boundaries. In consequence, a language such as English, where gender distinctions are carried solely by pronouns, is treated as a (natural) gender language. A contrasting view is represented by Wiese (1983: 373) who excludes pronominal anaphora from the local domain of agreement and from agreement boundaries. Nonetheless, the widespread approach has long been in favour of including anaphora, and this has been so for a reason, as is explained by Bresnan (1986: 278):

Grammatical agreement systems evolve historically from the morphological incorporation of pronouns into their governing predicates: for example, a subject pronoun becomes cliticized and then morphologically bound to its verb, and what begins as an anaphoric relation between the incorporated subject pronoun and a discourse topic somehow evolves into a grammatical agreement relation between a verb and its subject argument.

Although this approach has been criticized, and despite a few arising questions, such as the exact process of this change from a pragmatic to a syntactic relation, it shows that pronouns may have been in some part a source of grammatical agreement. In their in-depth discussion on the subject, Barlow and Ferguson (1988) agree that there is no reason for excluding anaphoric pronouns, and more recently, Corbett (2001: 4) gives two types of evidence supporting this conclusion: the features involved, and the distribution of syntactic and semantic agreement. In his work on gender, Corbett (1991: 112) claims that “attributive modifiers and pronouns are linked as poles of a single hierarchy, suggesting that they should be treated as parts of the same phenomenon”. This leads us to his Agreement Hierarchy (Corbett, 1991: 225), which has been designed to resolve instances of problematic gender agreement, such as those with hybrid nouns and resolution. Four types of agreement targets constitute the Agreement Hierarchy:

attributive < predicate < relative pronoun < personal pronoun

The following constraints apply to possible agreement patterns (Corbett, 2001: 7):
For any controller that permits alternative agreement forms, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement with greater semantic justification will increase monotonically (that is, with no intervening decrease).

Polish *dziewczę* 'girl(N)' is an example of a hybrid noun, which takes neuter agreement as one possibility, and feminine agreement as another. The former option is a syntactic "choice" resulting from the neuter declensional pattern to which this noun formally belongs. The feminine agreement can be chosen based on semantics. Let us consider the noun *dziewczę* in a given syntactic context:

(1) Widzia-ł-eś to ład-n-e dziewc-zą, ktor-e tam stoi?
    see-PST-2SG this.N pretty-N girl(N)-ACC who-N there stand-PRS.3SG
    *Jak-a ona eleganck-a!*
    how-F she smart-F

'Did you see this pretty girl who is standing there? How smart she is!'

The choice of agreement is resolved in terms of the Agreement Hierarchy and depends not only on the type of the target, but also on the distance from the controller: the attributive adjective *ładne* in (1) must take syntactic agreement, while the likelihood of the semantic agreement increases as we move rightwards to the personal pronoun *ona*.

As to the restrictions on gender agreement discussed by Corbett (1991: 122-135), they can occur in syntax, phonology and lexicon, and they may result from certain circumstances in which a particular word-class does not agree in gender. As an example on restrictions in syntax, Corbett (1991: 122) compares German adjectives, which show gender agreement only with their head noun; never in the predicate position. There can also be interactions with such categories as tense (gender agreement on verbs occurs only in the past tense), person (agreement depends on the person of the subject), number (agreement depends on the number of the subject and in some languages agreement is restricted to e.g., singular) and case (agreement depends on the case of the subject noun). An additional source of interaction between gender and case are subgenders, which will be discussed in more detail in §1.1.3. Another type of restriction can be found in the phonological system, as in French, where "phonological conditions produce a situation in which gender is not differentiated" (Corbett, 1991: 134). As to lexical limitations, the situation is less regular than in the case of other types of restrictions, as in Russian, where higher numerals do not agree in gender (Corbett, 1991: 135).

³ No hybrid nouns allowing both masculine and feminine agreement have been found in the children's data, and consequently, no analyses have been provided.
1.0.3 Gender assignment

A considerable amount of research has been done in the area of gender assignment since Bloomfield’s (1933: 280) pessimistic view that there are no criteria in a language that help determine the gender of a noun. Although there exist more and earlier studies dealing with gender assignment, the assignment systems presented here will be discussed on the basis of two studies dealing with world languages and their noun classes, namely Corbett’s (1991) detailed study on gender and Aikhenvald’s (2000) research in noun categorization devices. An additional source of information will also be Aikhenvald (2004). To begin, nouns are assigned to different genders (or noun classes') according to two sets of principles: semantic and formal (Corbett, 1991: 7-69). Any world language which distinguishes gender, will have some assignment system.

**GENDER ASSIGNMENT PRINCIPLES**

\[ \text{semantic} \quad \text{formal} \]

Formal principles can be further divided into:

\[ \text{morphological} \quad \text{phonological} \]

1.0.3.1 Semantic criteria

There are no purely formal gender assignment systems, i.e. all gender assignment systems contain semantic assignment rules. This rule of the ever-present semantic core, introduced by Aksenov (1984), is later paraphrased by Aikhenvald (2000: 21):

> There is always some semantic basis to the grouping of nouns into classes (genders), but languages vary in how much transparency there is. This semantic basis usually includes animacy, humanness and sex, and sometimes also shape and size.

The primary rule in semantic assignment systems is the ability to determine the gender of a noun by only knowing its meaning. Corbett (1991: 8-12) goes further in his analysis by ordering semantic gender systems on a scale from *strictly* semantic, through *predominantly* systems to *partially* semantic gender assignment systems. In strict semantic

---

1 Aikhenvald points to the fact that the three terms: noun class, gender and gender class are sometimes used interchangeably, and the choice depends on the linguistic tradition. She chooses to use noun class as a cover term for both noun class and gender, the latter of which she applies to small systems of two or three distinctions typical for Indo-European family (2000). In her latter work, however, Aikhenvald decides to use the term gender to avoid confusion (Aikhenvald, A. Y 2004). Corbett (1991) on the other hand, decides to use the term gender irrespectively of the language family.
assignment system, such as Tamil, all nouns are divided into two classes: rational (humans, gods, demons) and non-rational (others), the former being further divided into masculine and feminine (male gods/humans are masculine, goddesses/female humans are feminine; the residue nouns are neuter). Predominantly semantic assignment systems, e.g., Dyirbal or Zande, have so-called “leaks” containing nouns that belong to the semantic residue, i.e., a group of nouns not assigned by a semantic principle, the assignment of which can be somewhat random (Corbett, 1991: 13). Further towards the formal assignment systems are languages with partially semantic principles, such as Lak or other Daghestanian languages, where “the degree of semantic motivation varies from language to language” (Aikhenvald, 2000: 23). In Lak, there are relatively clear assignment rules for humans but less so for others: living beings are assigned to gender according to their sex (genders I and II), the non-rational animates and inanimates belong to gender III, while the residue to gender IV. In his analysis of the semantic assignment systems Corbett (1991: 32) summarizes this subject in the following way:

Given that apparently unrelated criteria may assign nouns to the same gender, it is tempting, when nouns do not fit into a semantic assignment system, to look for more and more semantic criteria which would account for them. In some languages, however, it soon becomes evident that there is instead a formal, rather than semantic criterion which will account for some or all of the nouns in the semantic residue.

Here in Corbett’s use, the term “semantic” refers to the natural gender (or perceived sex) principle (male human beings are assigned masculine, female human beings feminine). This is the most obvious semantic basis for gender distribution. Corbett’s description of assignment systems gives a wide-angled comparative view of gender systems in world languages. However, the literature also shows that there is a different way of looking at semantic assignment: more locally and within one language. Probably the most well known in this area is work done by Zubin and Köpcke (1984a), who analysed the gender distribution in the German vocabulary and found a highly structured gender system. A more recent study by Schwichtenberg (2004), which deals with semantic regularities not based on natural sex distinctions, suggest that semantic regularities might be part of gender assignment system of native speakers. In her view, contrary to categorical assignment rules and exceptional gender assignment, which are well investigated, “assignment regularities, i.e., tendencies in the gender distribution identified within the vocabulary of a language, are still controversial” (2004: 326). She investigates possible regularities in semantic assignment in German by presenting participants with a category (e.g. predator) and a pair of gender-marked nonce words. The participants choose a gender-marked determiner to go with the
nonce word, thus deciding on its gender. Schwichtenberg's analysis confirms observations made by other linguists as to the presence of semantic basis for the classification of nouns into classes (genders).

1.0.3.2 Morphological criteria

a. Inflection
Morphological and phonological principles apply only when semantic principles fail to provide a solid basis for assigning gender to a given noun. This happens largely in the Indo-European language family and Bantu languages, and it involves nouns in the "semantic residue" defined in the preceding section. Morphological assignment systems differ from phonological ones in that they require access to information about the declensional system of a noun (e.g., the form of the nouns in nominative and accusative), rather than just a single piece of information about which final vowel a noun has. There may be, however, some correlation between phonology and a declensional type. In Russian, for instance, nouns in declension III have palatalized stems (Corbett, 1991: 36). How exactly do the morphological rules operate? They assign gender to nouns in the semantic residue by "classifying" them according to their declensional patterns. In Russian, nouns of declensional type I are masculine, those of type II and III are feminine, while others are neuter (Corbett, 1991: 36).

The so-called hybrid nouns, i.e. nouns that display a conflict between different rules of gender assignment (Aikhenvald, 2004: 1033), pose another problem for the morphological assignment principles, mainly because in the case of hybrids the morphological and semantic rules clash. Semantic criteria cannot take over, since such nouns can denote both a female and a male being as in the case of the Polish term prezydent ('head of the state'). This noun does not simply belong to both feminine and masculine gender. Following Corbett (1991: 225): "The crucial point about hybrid nouns is that the form of gender agreement used with them depends in part on the type of agreement target involved". In Polish as well as in Russian, morphological assignment criteria do not cover masculine nouns ending with the vowel -a. They are not excessively numerous, but they may surprise a foreigner learning any of those languages and trying to solve the mystery of gender assignment. The inflectional class which has the ending -a for the singular nominative case is the one to which nouns denoting females are typically assigned, but some nouns which denote males, may be assigned to this class. These nouns are for example atleta, poeta or mężczyzna. We will

\[5\] "...if in order to establish the gender of a noun we need to refer to more than one form ... then we are dealing with a morphological assignment rule. If, on the other hand, gender can be established by reference to a single form, then we are dealing with a phonological rule." (Corbett 1991: 51).

\[6\] More details on -a ending and its occurrence may be found in § 1.1.2.

21
return to Polish in the next section of this chapter with a more detailed analysis of the features underlying morphological gender assignment systems.

b. Derivation

Derivational assignment rules can be based on such morphological processes as suffixation or derivation of deverbal nouns. Frequently cited examples of German diminutives in \textit{-chen}, \textit{-lein} are instances of suffixes determining the assignment of neuter gender to a noun. Less frequently one can find examples of prefixes, which would determine the choice of gender. Again, often cited example is German prefix \textit{Ge-} as in \textit{das Gemisch} 'mixture'.

1.0.3.3 Phonological criteria

A clear example of a phonological assignment system is observable in Qafar (East-Cushitic) (Corbett, 1991: 51), where the position of the accent is an indicator of gender. Thus, nouns whose citation form ends in an accented vowel are feminine, while all others are masculine. Another example is Katcha (Kordofanian) where any noun beginning with \textit{m-} is feminine (Aikhenvald, 2004: 1035). Clearly, the rule is based upon the stem, rather than on the type of declension, as it was in the case of morphological assignment systems.

Let us for a moment return to German. The work by Schwi cttenberg (2004) investigating semantic gender assignment based on regularities found in non-human nouns shows that in "non-associated categories" (neuter), phonological assignment criteria are preferred. Zubin and Köpcke (1983) also suggested that phonological criteria are preferred in gender assignment in forced-choice tasks. Since no linguistic system in the world is a purely formal system (based exclusively on morphological or phonological criteria or both) without the semantic core, it can be difficult to discern the type of assignment. There can occur numerous instances of nouns whose semantics, morphology and phonology indicate the same choice of gender. Schwi cttenberg (2004) asks: How do assignment regularities interact to provide the gender for a specific noun? In other words, what is the interplay between the formal and semantic factors? Some scholars claim that there is a strict hierarchy (Wegener, 1995), others that no strict hierarchy of rules is possible (Salmons, 1993). Wegener proposes a hierarchy that runs from morphological, then semantic, then phonological regularities. Salmons, on the other hand, believes that instead of a hierarchy we have "continua based on relative strength or weakness of particular tendency and the degree of membership in a particular semantic class that a particular word shows" (Salmons, 1993: 426).
1.1 Gender in Polish

1.1.1 The nature of gender in Polish

The distinctiveness of the evolution of the category of gender in Polish has been highlighted by Hjemslev (1956) in his analysis of the Slavonic gender systems. The *sui generis* nature of the Slavonic systems, as Hjemslev has shown, "is the result of the interplay of two contradictory tendencies in the evolution of the grammatical gender systems: the conservative tendency to sustain the inherited gender system and the tendency towards semantic motivation (semantic transparency) of grammatical gender distinctions" (after Laskowski, 1986: 459). Such a striking complexity of the gender system in Polish is both fascinating and puzzling. While English has lost its grammatical gender distinctions, Polish has introduced new subgenders. It seems to be hardly surprising that this multifaceted phenomenon has generated volumes of research. At a first view, the Polish gender system seems disorderly and perplexing. Research on its origin reveals the process of gradual weakening of the semantic motivation of gender distinctions, and the resulting grammaticalisation of the gender system. The addition of the new semantically motivated dimensions *personal-nonpersonal* and *animate-inanimate* to the earlier Indo-European system of masculine, feminine and neuter gender, made it even more complicated, especially with the masculine being the only gender to be the locus of these new features. Furthermore, while the masculine gender has undergone deep structural changes, the neuter gender has retained its most important inherited inflectional features (Laskowski, 1986: 462), which led to a divergent development of the two genders. Over the past fifty years there has also been a heated debate on the number of genders in Modern Polish. It has been shown that Polish grammatical gender can have either three, five or nine genders, depending on the analytical criteria applied. In addition, any changes in the system resulting from the users' everyday choices, as well as dialectal and idiolectal modifications, have led to the Polish gender system becoming remarkably complex to describe and analyse.

Gender in Modern Polish is marked on adjectives, pronouns, numerals and verbs in the past tense. It is the sole inherent feature of nouns, which are always inflected for case and number. Here we slowly approach a difficult and often confusing distinction between gender and the inflectional system. Both categories are always present on Polish nouns, and have been on many occasions confused by researchers trying to interpret gender through inflectional suffixes, and as a result, making it impossible for others to interpret their results. All nouns, apart from a few borrowings and foreign names, inflect for seven cases, although there is a growing tendency for substituting the Vocative with the Nominative. The inflectional patterns may at times seem complex due to various consonant and vowel
alternations involved in the process. With the exception of the two defective noun classes, i.e. 
singularia tantum (e.g., zło 'evil(N)') and pluralia tantum (e.g., spodnie 'trousers(N_M.PERS)'), a declensional paradigm for each noun consists of fourteen cells. Yet, the number of distinctive inflection suffixes attached to each noun is much smaller due to the 
syncretism of those suffixes. Thus, Polish nouns have a minimum of five (e.g., muzeum(N)) and a maximum of nine (e.g., niebo 'sky(N)') inflectional affixes in their paradigm. Rather 
typically for a rich inflectional system, Polish affixes combine the concepts of gender, case 
and number, and separating them is impossible. For this very reason, the gender of a noun 
cannot be specified based solely on the inflectional affix. As an example of the inflectional 
richness of Polish, let us consider the ending -a, which occurs in at least 6 different 
inflectional cells, i.e. has at least six different functions in Polish inflectional system:

(2) Siostr-a^ Michal-a zjadl-a ciast-o
Siostra(F)-NOM.SG Michal(M)-GEN.SG ate.PST.PFV-P.3S cake(N)-
ACC.SG
pod stol-em w kuchn-i
under table(M)-INS.SG in kitchen(F)-LOC.SG
'Michael's sister ate a cake under the table in the kitchen.'

(3) Natychmiast wezwa-no lekarz-a
immediately call-PASS.PST doctor(M)-ACC.SG

Example (2)^ illustrates two possible functions: nominative singular in the feminine pattern 
(siostra), and genitive singular in the masculine pattern (Michala), which is also illustrated in 
example (3), but this time it is the accusative singular (lekarza). More functions, this time 
involving the plural, are exemplified by (4) and (5):

(4) Nasz-e okn-a wychodz-q na podwórze
our-N_M.PERS.NOM.PL window(N-M.PERS)-NOM.PL face-PRS.3PL at yard(N)-ACC.SG

'Our windows face the yard.'

(5) Widz-q nasz-e poľ-a z daleka
see-PRS.1SG our-N_M.PERS.NOM.PL fields(N_M.PERS)-ACC.PL from distance

'I see our fields from a distance.'

The suffix -o functions as a plural marker of nominative in (4) (okna), and of accusative in 
(5) (pole), in both cases for the neuter pattern. Smoczyńska's list of the occurrences of the 
suffix -a in the neuter noun pattern fails to note one more instance, namely when suffix -a 
functions as a genitive singular. An example here could be (6):

7 Bolding mine for emphasis.
8 All examplea taken from Smoczyńska (1985b).
Experienced researchers are rarely surprised by exceptions they find as they study a linguistic system. The Polish gender system is not devoid of them. The nominative singular -a marker earlier exemplified in the feminine class (siostra) is also found in the masculine class. Such nouns are for example poeta 'poet(M)', artysta ‘artist(M)’, atleta ‘athlete(M)’. Although it is not an extensive group, it must be mentioned here, and perhaps even treated separately as another use for the -a affix. Masculine nouns with the -a affix serve a good purpose of illustrating the distinction between semantic and grammatical gender. A language has a semantic (natural) gender system whenever the gender of a noun can be predicted on the basis of its meaning, i.e. a noun denoting a female will be feminine, one denoting a male will be masculine. As has been mentioned earlier, semantic gender assignment systems can be strictly semantic, predominantly semantic or partially semantic. As to the grammatical gender system, it is present in a language whenever the gender of a noun triggers a modification of the morphological structure of all those parts of speech that pertain to that noun within one sentence or beyond. Polish adjectives, numerals, pronouns and verbs in the past tense receive their gender from nouns. Consequently, these parts of speech remain in agreement with the noun that they describe or relate to.

1.1.2 Gender assignment

1.2.2.1 Semantic criteria

Similarly to other semantic assignment systems, the notion of animacy plays the most important role in determining the gender of a noun on the basis of its meaning. According to the semantic assignment principles in Polish, male humans and higher male animals are assigned masculine gender, female humans and higher female animals are assigned feminine gender, while the non-sex-differentiable residue may be assigned either masculine, feminine or neuter gender based on their form. Below are some examples of semantic assignment rules in Polish:

MASCULINE: 


FEMININE: 

Exceptions to the semantic rules operating in Polish can be shown with the use of diminutives, augmentatives, young animates, and downgrading of nouns denoting humans. Consider the examples below:

(7) \( T\text{-}en\)  \( \text{maly}\)  \( \text{kot} \)  vs  \( T\text{-}o\)  \( \text{maly}\)  \( \text{koci\'g} \)

\( \text{this-M little-M cat(M)} \)  vs  \( \text{this-N little-N cat(N),DIM} \)

"This little cat."

The diminutive form of the noun \( \text{kot} \)  ‘cat(M)’ triggers off the change from masculine into neuter gender across all parts of speech in the example above, so \( \text{ten} \) becomes \( \text{to} \) and \( \text{maly} \) ‘little’ changes into \( \text{male} \). Similar transformations take place with the use of augmentatives. Note how the gender of the adjective \( \text{okropna} \) ‘horrid’ changes from feminine into neuter under the influence of the change of the feminine noun \( \text{baba} \) ‘woman’ into an augmentative neuter noun \( \text{babsko} \):

(8) \( C\text{o} \text{ za} \)  \( \text{okropn-a} \)  \( \text{baba} \)  vs  \( C\text{o} \text{ za} \)  \( \text{okropn-e} \)  \( \text{babsko}! \)

\( \text{what.kind.of horrid-F woman(F)} \)  vs  \( \text{what.kind.of horrid-N woman(N),AUG} \)

"What a horrid woman!"

The semantic gender rule always takes over when it comes to the earlier mentioned group of masculine nouns that (rather atypically) end in \(-a\). Let us now turn to the so-called semantic residue, which includes a very extensive group of inanimate nouns, for which no natural gender rule can be applied. According to Corbett (1991: 35), in Russian it is impossible to establish semantic factors to account for the gender of non-sex-differentiable nouns making up the semantic residue. A similar conclusion can be drawn with reference to Polish, where these nouns are distributed to masculine, feminine and neuter gender either arbitrarily, if we refer to semantics and follow Trudgill (1999), or non-arbitrarily, if we refer to formal principles of their assignment, such as their declensional class. A comparison of Polish and Russian shows how nouns which belong to two different declensional classes but have identical meaning and almost identical spelling are assigned to different genders: Russian \( \text{flag} \) ‘flag’ is masculine, while Polish \( \text{flaga} \) ‘flag(F)’ is feminine. The formal factors, which allow Polish residue nouns belong to one of the three main genders, are predominantly of the morphological type. An extensive overlap of semantic and formal criteria is observable in Polish.

26
1.1.2.2 Formal criteria

Formal criteria of gender assignment, as opposed to semantic criteria, include morphological and phonological properties of each noun. As mentioned in §1.0.3 above, although there may be some correlation between morphological and phonological clues, morphological assignment differs from phonological in that it requires information available only from the declensional system of a noun. It has already been explained that no language has a purely formal system (Corbett, 1991: 34), and Polish is no exception in this respect. The semantic criteria apply first, and when those fail to provide enough information about which gender the noun should receive, then either morphology or phonology becomes the source of information. Semantic criteria may overlap with the formal criteria, e.g., the noun *mama* 'mum' denotes a female, but it also belongs to a declensional class that typically assigns feminine gender. Semantics and form may also be in conflict with each other, as in for example *tata* 'dad', which is semantically masculine, but belongs to the same declensional class as *mama*. The source of the conflict is undoubtedly clear.

When discussing formal criteria of gender assignment in Polish it is easy to fall into the already mentioned trap of confusing gender and inflection. Materials published on the Polish gender system often fail to clearly differentiate between the two. To clarify, declensional classes are recognised by their patterns. Based on the minimal number of declensional suffixes (usually two), we can identify a declensional class of a given noun. Genders, on the other hand, are identified based on agreement between controllers (nouns) and targets (adjectives, numerals, determiners). When the two sources of information about the noun are put together, it becomes clear that nouns whose gender is identified via agreement fall into particular declensional classes. Animate nouns whose gender is identifiable from the natural gender rule do not come into question when discussing the principles of the formal gender assignment. For this, inanimate nouns are the focus. Since there is no underlying semantic rule for nouns such as *stôl* 'table(M)' and *torba* 'bag(F)', in terms of semantic gender these two nouns are indistinguishable. Their gender is recognised through agreement between them and parts of speech referring to them. Moreover, apart from single-gendered nouns, there are also multi-gendered nouns in Polish, which take a

---

9 "Formal semantic systems are really semantic plus formal systems" (Corbett 1991: 308).
10 In the study of gender in German, Schwichtenberg and Schiller (2004) found that when "phonological and semantic assignment regularities conflicted...both assignment regularities worked together such that neither determiner was preferred". In other words, a similar number of nouns with conflicting regularities have been allocated to two different gender classes, and neither semantic nor formal criteria were preferred. Interestingly, for words marked with the masculine marker -er in feminine group, a preference for the semantically motivated gender has been observed, suggesting that semantic criteria overrule formal criteria regardless of whether the formal criteria point to masculine or feminine gender.
different gender depending on the context and the choice of the referent made by the speaker. Lamaga ‘fumbler’ is typically regarded as a hybrid noun, i.e. one that takes a different gender with different targets in one utterance. It can also serve as an example of a multi-gendered noun also present in English. An example for English is the noun baby.

When we look at the Polish noun inventory, some single-gendered nouns have gender assigned to them according to semantic criteria (animate nouns), and some according to form, i.e. inflectional class (inanimate nouns in particular). Although semantic rules in Polish operate with a similar accuracy as in Russian, i.e. they “operate with very few exceptions” (Corbett, 1991: 34), they only cover semantically distinguishable nouns. The remaining nouns, i.e. those whose gender is not semantically distinguishable, are referred to as “semantic residue”. In other words, those nouns are semantically sexless, or simply inanimate. Gender for those nouns is determined not by semantics, but by form, i.e. the phonological endings, and by the declensional types. Phonological criteria help identify gender, but they are not always successful. According to phonological assignment rules in Polish, nouns ending in a consonant are masculine; nouns ending in the vowel -a are feminine; those ending in -ç, -o, and -um neuter. There are also exceptions: a small group of masculine nouns ends in -a, e.g., artysta ‘artist(M)’, and some feminine nouns end in a consonant, e.g., krew ‘blood(F)’. Establishing their gender is possible based on the declensional class of those nouns. The nominative forms alone would not be sufficient due to the overlapping marking. Determining the gender requires “access to more than one case form of the noun, in other words, to its declensional type” (Corbett, 1991: 36). As an example, let us take the noun os ‘axis(M)’. In Polish, many masculine and feminine nouns end with a soft consonant, thus phonological properties of os do not point to a particular declensional class. We must refer to the morphological shape of this noun, i.e. its inflectional pattern in either the genitive or the instrumental singular: Nom. os, Gen. osi, Inst. osiq. Knowing the nominative form plus one of the two remaining cases would be sufficient to conclude that the os belongs to the feminine declensional pattern. Neither masculine nor neuter has the -i suffix in the singular genitive or the -q suffix in the instrumental. Since the morphological assignment rules require access to more than one case form of the noun, i.e., its declensional type, we now turn to declensional patterns in Polish. Regrettably, linguists have failed to provide a reliable and, above all, a united description of the Polish declensional class system. The classification of Polish declensional classes is far from clear, as each description seems to be relying heavily on how each author categorizes various formal features of the Polish noun. The classic textbook on Polish morphology by Grzegorczykowa (1998) provides a lengthy and detailed analysis of the Polish noun. Any hope to find a clear statement on the number of Polish declensional classes disappears as
soon as we open the section on nouns. Instead, this important textbook classifies nouns according to genders for a start, which is then extended into further subclasses. The author seems to be trying to identify all paradigmatic differences in the declensional system, thus failing to provide any information on how those classes can be handled in a more precise and practical way. The extract below shows how unnecessarily detailed, hence impractical, the noun classification in this important textbook is:

Deklinacja żeńska (Feminine declension):
1. Klasy deklinacyjne leksemów z końcówką M. lp -a (Decl. class of lexemes with -a ending in Nom Sg):
   1.1. Podklasa leksemów twardotematowych (Subclass of hard-stem lexemes):
      1.1.1. Grupa paradygmatyczna z cechami dystynkcyjnymi (Paradigm. group with distinctive features): M. lp -a-zero, M Im -y/-i-zero, D. Im -zero — np. ryba, łąka
      1.1.2. Grupa paradygmatyczna z cechami dystynkcyjnymi (Paradigm. group with distinctive features): M. lp -a, M. Im -e, D Im -zero — np. szansa

For the sake of clarity, based on Corbett’s classification of Russian noun declension (Corbett, 1991: 36), we shall classify Polish nouns into four main noun paradigms:

<table>
<thead>
<tr>
<th>Singular</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>stól (‘table’)</td>
<td>szkoła (‘school’)</td>
<td>kość (‘bone’)</td>
<td>wino (‘wine’)</td>
</tr>
<tr>
<td>Accusative</td>
<td>stół</td>
<td>szkołę</td>
<td>kości</td>
<td>wino</td>
</tr>
<tr>
<td>Genitive</td>
<td>stołu</td>
<td>szkoły</td>
<td>kości</td>
<td>wina</td>
</tr>
<tr>
<td>Dative</td>
<td>stołow</td>
<td>szkołom</td>
<td>kośćmi</td>
<td>wina</td>
</tr>
<tr>
<td>Instrumental</td>
<td>stołem</td>
<td>szkołami</td>
<td>kośćmi</td>
<td>winami</td>
</tr>
<tr>
<td>Locative</td>
<td>stoach</td>
<td>szkolach</td>
<td>kościach</td>
<td>winach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plural</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>stolicy</td>
<td>szkoły</td>
<td>kości</td>
<td>wina</td>
</tr>
<tr>
<td>Accusative</td>
<td>stolę</td>
<td>szkoły</td>
<td>kości</td>
<td>wina</td>
</tr>
<tr>
<td>Genitive</td>
<td>stoliów</td>
<td>szkołół</td>
<td>kości</td>
<td>win</td>
</tr>
<tr>
<td>Dative</td>
<td>stolom</td>
<td>szkołom</td>
<td>kościami</td>
<td>winom</td>
</tr>
<tr>
<td>Instrumental</td>
<td>stołami</td>
<td>szkolami</td>
<td>kościami</td>
<td>winami</td>
</tr>
<tr>
<td>Locative</td>
<td>stóchach</td>
<td>szkolach</td>
<td>kościchach</td>
<td>winach</td>
</tr>
</tbody>
</table>

Corbett’s classification accommodates the nouns from the semantic residue, i.e. the morphological assignment rules involve inanimate nouns only. Corbett’s paradigm does not involve animate nouns (including masculine personal nouns), which in the plural form the masculine personal gender class. The non-masculine personal group involves all other nouns. The classifying elements can be observed via agreement:

11 Feminine nouns ending in -i such as bogini ‘goddess’ fit into declension pattern III, whereas neuter nouns such as stiele ‘herb’ or morze ‘sea’ fit into the declension pattern IV.
Let us now examine on what basis researchers have tried to establish the number of genders in Polish. This subject is inherently related to declensional classes, and thus seems to be the natural step to follow in the attempt to provide a fuller picture of declensional class and genders in Polish.

### 1.1.3 How many genders in Polish?

In many languages, the number of genders is quite indisputable, whilst in others it has been a source of heated debates for many years. How do we measure the number of genders in a language? Do those measurements have clear principles? The number of genders in Polish can still be regarded as an open question, since various scholars have applied different criteria. Męczak (1956) argues that Polish nouns can be divided into five genders, and his approach has been supported by Laskowski (1984) and Saloni-Swidzihski (1985). Other linguists claim that Polish has more genders: according to Corbett (1983) there are six, while Wertz (1977) finds seven.

Corbett (1991: 4) explains that “while nouns can be classified in various ways, only one type of classification counts as a gender system; it is one which is reflected beyond the nouns themselves”. In other words, the only way to attest how many genders a language possesses is to examine how many agreement combinations exist between nouns and other word classes in a given language. Agreement, as explained in the preceding section, is the essence of gender. No other means can serve as a determining criterion for gender, since, as Hockett (1958: 321) rightly stresses, “gender is reflected in the behaviour of associate words”. In Polish, nouns take agreement with the following parts of speech:

<table>
<thead>
<tr>
<th>Plural</th>
<th>Masculine personal</th>
<th>Non-masculine personal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>el panowie/lekarze/chłopcy</td>
<td>te stoły/szkoły/kości/wina</td>
</tr>
<tr>
<td>Accusative</td>
<td>tych panów/lekarzy/chłopców</td>
<td>te stoły/szkoły/kości/wina</td>
</tr>
<tr>
<td>Genitive</td>
<td>tych panów/lekarzy/chłopców</td>
<td>tych stołów/szkół/kości/win</td>
</tr>
<tr>
<td>Dative</td>
<td>tym panom/lekarzom/chłopcom</td>
<td>tym stolem/szkołom/kościami/winami</td>
</tr>
<tr>
<td>Instrumental</td>
<td>tymi panami/lekarzami/chłopcami</td>
<td>tymi stolem/szkołami/kościami/winami</td>
</tr>
<tr>
<td>Locative</td>
<td>tych panach/lekarzach/chłopcach</td>
<td>tych stolech/szkołach/kościami/winach</td>
</tr>
</tbody>
</table>

(These 'men'/‘doctors'/‘boys' (these 'tables'/‘schools'/‘boses'/‘wines')
Table 1. Elements taking agreement in Polish

<table>
<thead>
<tr>
<th>Parts of speech</th>
<th>Masculine</th>
<th>Feminine</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td>adjectives</td>
<td>nowy dom</td>
<td>nowa szafa</td>
<td>nowe krzeslo</td>
</tr>
<tr>
<td>pronouns</td>
<td>mój dom</td>
<td>moja szafa</td>
<td>moje krzeslo</td>
</tr>
<tr>
<td>numerals</td>
<td>pierwszy dom</td>
<td>pierwsza szafa</td>
<td>pierwsze krzeslo</td>
</tr>
<tr>
<td>participles</td>
<td>pomalowany dom</td>
<td>pomalowana szafa</td>
<td>pomalowane krzeslo</td>
</tr>
<tr>
<td>verbs (past tense)</td>
<td>dom stal</td>
<td>szafa stała</td>
<td>krzeslo stało</td>
</tr>
</tbody>
</table>

It is quite simple to reconstruct three sentences which would be built with as many parts of speech given above as possible. Note how all the parts of speech agree with the head noun from which they take their genders:

(9) **Mój pierwszy nowy DOM nie był pomalowany i stał nad rzeką.**
    'My first new house wasn’t painted and stood on the bank of a river'

(10) **Moja pierwsza nowa SZAVA stała w moim pokoju i była pomalowana na biało.**
     'My first new wardrobe stood in my room and was painted white'

(11) **Moje pierwsze nowe KRZESLO zostało pomalowane na zielono i stało przy biurku.**
     'My first new chair was painted green and stood by the desk'

The three head nouns in ex. (9)-(11) are used in the nominative singular. On this basis, Klemensiewicz (1965: 51) distinguishes three genders in Polish: masculine, feminine and neuter. This approach is understandable for two reasons: firstly, such is the tradition of treating the Polish gender system\(^{12}\), and secondly, nominative singular is the form most often referred to and it is the most basic one too. The works of other scholars show that they all agreed that such an approach leads to a oversimplification of the issue in question.

Mańczak (1956) argues that, depending on their form in the accusative, Polish nouns can be divided into five genders. His approach can be briefly described as follows. The gender of the noun cannot be judged based on the endings that head nouns have, as nouns that have the same type of ending often belong to distinct genders (e.g., *poeta* ‘poet(M)’ vs. *kobieta* ‘woman(F)). Mańczak (1956: 118) also stresses the fact that “w

---
\(^{12}\) According to Mańczak Mańczak, W. (1956). Ile rodzajów jest w polskim? Język Polski, 36, 116-121., this “tradition” should rather be regarded as a source of many inconsistencies resulting in the inconsequent argumentation about the number of genders in Polish. He claims that “Ponieważ w grece i łacinie są tylko trzy rodzaje, językobrazujący w ramach gramatyki klasycznej języki nowożytne staram się nie wykraczać poza tę liczbę” [Since Greek and Latin have only three genders, linguists have been trying not to exceed this number by squeezing grammars of modern languages into the frame of Latin and Greek grammar - translation mine]
gramatyce (polskiej przyp.) obowiązuje nie wyrażana zazwyczaj expressis verbis zasada ustalania liczby kategorii fleksyjnych w oparciu o największe zróżnicowanie paradigmatów. The rule to which Mańczak refers as the non-expressis verbis one is based on the assumption that the biggest number of paradigmatic distinctions should be the ground for positing any new inflectional categories. In other words, for the category of gender in Polish it means the largest number of distinctions of the target gender. The easiest choice to illustrate the number of those distinctions in the target gender is to look at how adjectives change their shape when used in the accusative. Mańczak uses accusative as the basis for positing new genders. Below is a graphic representation of Polish adjectives as targets, and their forms in the accusative in both the singular and plural:

**SINGULAR ACC.**
- mężczyzna 'man'
- dobrego 'good'
- psa 'dog'
- dobry
- stół 'table'
- dobrą
- dobrym

**PLURAL ACC.**
- dobrych mężczyzn
- dobrych kobiet
- dobrych stół
- dobrych dzieci

Mańczak does not deny that there are four genders in the singular and two in the plural in Polish (four agreeing adjective forms in the singular and two in the plural), but stresses that it is of the utmost importance to be consistent in describing the system (i.e. using the same criteria). If we agree that there are seven cases in the singular and six in the plural, we can say that there are four genders in the singular and two in the plural. However, if we claim that Polish nouns decline in seven cases with the syncretism in nominative and vocative, we are forced to state that Polish adjectives agree in five genders with the following kinds of syncretism: masculine personal and animal in the singular, and masculine animal, masculine inanimate, feminine and neuter in the plural (Mańczak, 1956: 120).

The theory presented by Mańczak has also been supported by Saloni (1976) and Rothstein (1976). The latter says about Mańczak:

\[13\] There has always existed a rule in the Polish grammar, which has never been expressed in a clear way that the number of inflectional categories should be established on the basis of the most numerous paradigms [translation mine].
Saloni has also posited the existence of nine agreement classes based on the agreement between nouns and adjectives, verbs but also numerals. With his approach, a question arises as to which parts of speech are more central, and what “more central” should mean. Rothstein (1976: 249) seems to give sufficient evidence that in agreement \textit{ad sensum}, numerals in Polish can vary in form quite considerably, especially with nouns of mixed agreement (in ex. 12-14 the intended case in NOM):

\begin{itemize}
  \item (12) \textit{kilku} bandziorów/dryblasów/liszusów/starych pryków \textit{a few} bandits/strappers/boot-lickers/old farts
  \item (13) \textit{dwoch} drabów/obiboków \textit{two} bandits/layabouts
  \item (14) \textit{czterech} obdartusiów \textit{four} crusties
\end{itemize}

Regular nominative forms \textit{dwa\textasciitilde{}, trzej, cztery\textasciitilde{}} do not exist in connection with those nouns. However, infrequent exceptions take place: \textit{dwa pijusy} ‘two guzzlers’. It is possible that agreement rules that govern numerals in Polish are “less strict” and “less regular” than those governing adjectives or verbs, and as a result, some linguists consider them “less central”. Brooks and Nalibov (Brooks and Nalibov, 1970), as well as Schenker (1964) represent a different approach from the one followed by Mańczak. In their opinion, the division between the singular and plural is vital for Polish, for the reason explained by Schenker (1964: 15):

Mańczak contends that it is inconsistent to posit one set of gender distinctions in the singular and another one in the plural without doing it at the same time for other grammatical categories in analogous situations. He points out that the Latin plural, in contrast to the singular, shows constant syncretism between the dat. and abl....However, the instances quoted by Mańczak are not comparable with gender in Polish since \textit{none of the gender categories distinguished in one number has an exact counterpart in the other} [emphasis mine]

In his earlier work on Polish gender, Schenker (1955) developed the idea of “distinctive environments” originating from the examination of different paradigm forms in Polish, which reveal a correlation between a difference in the shapes of forms and difference

\footnote{14 It has become a truism to claim that there are three genders in the singular in Polish...and in the plural there two, but different from the singular ones. Witold Mańczak broke off this tradition and suggested a five-gender system...including entire noun paradigms, instead of only the forms of the singular and the plural [translation mine].}

\footnote{15 \textit{ad sensum} - semantic agreement vs. \textit{ad formam} - grammatical agreement}
of their environment. In other words, the presence of some endings triggers the appearance of others. Since Schenker’s work was published a year earlier, it might be more appropriate to say that it was Mańczak who supported Schenker’s view. In any case, both linguists agree to the idea of “greatest diversity of forms” as the basis for analysis. Schenker (1955: 402) defines his distinctive environment as “an environment which determines the function of a given inflectional category denoted by an inflectional ending is called the distinctive environment of that category”. The main prerequisite in his theory is that the number of genders in Polish should be based on case and number, rather than number alone. This approach leads Schenker to posit seven genders on the basis of the agreement of adjectives with nouns, which he claims “exhibit the greatest differentiation of gender forms” and thus, adjectival paradigm “is the logical starting point in determining gender categories” (Schenker, 1955: 403). Table 2 below illustrates the adjectival forms of Polish (vocative is not included). Similarly to Mańczak, Schenker’s paradigms distinguish four adjectival endings based on the accusative singular. In Schenker’s words, “in the environment of the accusative four distinct forms occur”, while in the environment of nominative there are three gender categories. In the environments of the remaining cases, only two gender categories emerge. The way Schenker establishes that the number of gender categories in the singular is different from the one applied in the plural. In the singular, he establishes as many categories in each case as there are distinct forms, while in the plural, five distinct categories are established based on the accusative and nominative forms (Table 3)\textsuperscript{16}.

<table>
<thead>
<tr>
<th>Table 2. Adjectival paradigms in Polish – singular (Schenker, 1955)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N. Sg.</strong></td>
</tr>
<tr>
<td><em>noga</em> ‘leg’</td>
</tr>
<tr>
<td>jedna</td>
</tr>
<tr>
<td>jednej</td>
</tr>
<tr>
<td>jednej</td>
</tr>
<tr>
<td>jednej</td>
</tr>
<tr>
<td>jednej</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Adjectival paradigms in Polish – plural (Schenker, 1955)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N. Pl.</strong></td>
</tr>
<tr>
<td><em>noga</em> ‘leg’</td>
</tr>
<tr>
<td>dwie</td>
</tr>
</tbody>
</table>

\textsuperscript{16} Numeral adjectives *jedn-* ‘one’ and *dw-* ‘two’ are used in Table 3 and Table 4, since they “present the largest number of distinct forms” (Schenker, 1955: 403).

\textsuperscript{17} I have taken the liberty of exchanging Schenker’s outdated spellings, such as “ow” into “q” and “jednim” into “jedny”.
Schenker's distinctive environment theory necessitates positing such genders as masculine special and masculine depersonalised based on their distinct forms, one in the nominative, the other in the accusative. In total, there are nine sets according to their occurrence as modifiers of different nouns. Schenker (1955: 405) also shows that two new genders have to be established on the basis of the environment of the genitive plural (feminine + nonfeminine). It must be noted that the majority of Schenker's examples are nouns such as przybîdła 'stray', gaduia 'chatterbox', modnis 'dandy', kaleka 'cripple', niezdara 'butterfingers', beksia 'crybaby'. Rothstein (1993: 697) treats these nouns as exceptions, and in fact, they all have a choice of agreement patterns, and therefore, should not be treated as the source of new genders in Polish.

The last opinion to which we shall turn now is the one represented by Corbett (1983, 1991), who rejects the widely supported theory of a quintuple gender system, as well as Wertz's theory of seven genders, on the assumption that Wertz's seven-gender system is unable to handle the surface facts of agreement in Polish (1983: 83). The cornerstone for the entire debate is the question whether gender in the singular should be treated as separate from that in the plural, or not. Schenker (1964) and Wertz (1977) support the split of genders systems in the singular and plural, on which basis they arrive at a different number of genders in Polish. Corbett (1983: 86) rejects the necessity of separating singular and plural based on the lack of matching categories in the two, and suggests the six-gender system: masculine personal, masculine devirilised\(^\text{18}\), masculine animate, masculine inanimate, feminine and neuter. In his later work, Corbett (1991), deals with the rise of subgenders in Slavonic languages in a more detailed way, claiming that "a major factor in this development was the requirement to distinguish subject from object". The nouns involved were originally specific male humans, and other masculine nouns followed later on (at the same time, weakening the semantic conditions of the subgender by inanimates behaving as animates) (Corbett, 1991: 99). Since for most masculine nouns, the morphological distinction between nominative and accusative has been lost in the singular, the use of genitive for accusative allowed this distinction. According to Corbett, the inanimates that were treated as animates (e.g., pech 'misfortune(M)') "may (…) serve as Trojan horses for the final loss of the animacy distinction" (1991: 99). Furthermore, "the existing markers for agreements in case were reallocated...to give new agreement classes, which are the basis for the new subgenders" (Corbett, 1991: 313).

To count genders, Corbett (1991) advises determining which types of agreement class should be recognised as genders, so that not more agreement classes are identified than

\(^{18}\) Corbett's masculine devirilised gender (e.g., karzel 'midget') is based on the same pattern as Schenker's masculine depersonalized gender (e.g., cham 'yob').
it is “intuitively satisfying” for a number of genders in a given language” (1991: 145). Remembering that the agreement class approach (§1.0.2.) involves the notions of target and controller genders, we shall refer again to those parts of speech (target genders), which show agreement in Polish. They are: adjectives, verbs, pronouns, participles and numerals. Indeed, all the possible nouns in Polish together with all their possible contexts of use would lead to identifying more agreement classes than we would intuitively be able to agree to accept. Let us now define the term subgender and on this basis see how Polish fulfils the requirements described in Corbett (1991: 161-8) with reference to Serbo-Croat and Russian:

Subgenders are agreement classes which control minimally different sets of agreement, that is, agreements differing for at most a small proportion of the morphosyntactic forms of any of the agreement targets.

Similarly to the notion of “more central” and “less central” parts of speech mentioned earlier, the idea of “small proportion” is a matter of judgement. The data below follows Corbett’s instructions of listing all the separate agreement requirements for members of different agreement classes and comparing the results: Corbett says: “If comparison of the difference between two agreement classes yields no more than a minimal difference (compared with the general level of differences between other agreement classes) then the two agreement classes in question are subgenders” (1991: 163-4). Consider the data on attributive agreement shown in Table 4 (vocative is included for comparative purposes):

<table>
<thead>
<tr>
<th>Singular</th>
<th>ten student</th>
<th>ten kot</th>
<th>ten stol</th>
<th>ta szkoła</th>
<th>to wino</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominative</td>
<td>‘this student’</td>
<td>‘this cat’</td>
<td>‘this table’</td>
<td>‘this school’</td>
<td>‘this wine’</td>
</tr>
<tr>
<td>accusative</td>
<td>ten student</td>
<td>ten kot</td>
<td>ten stol</td>
<td>ta szkoła</td>
<td>to wino</td>
</tr>
<tr>
<td>genitive</td>
<td>tego studenta</td>
<td>tego kota</td>
<td>ten stol</td>
<td>tej szkoły</td>
<td>tego wina</td>
</tr>
<tr>
<td>dative</td>
<td>temu studentowi</td>
<td>temu kota</td>
<td>temu stolowi</td>
<td>tej szkole</td>
<td>temu winu</td>
</tr>
<tr>
<td>instrumental</td>
<td>tym studentem</td>
<td>tym kota</td>
<td>tym stolem</td>
<td>tą szkołą</td>
<td>tym winem</td>
</tr>
<tr>
<td>locative</td>
<td>tym studentce</td>
<td>tym kocie</td>
<td>tym stole</td>
<td>tej szkole</td>
<td>tym winie</td>
</tr>
<tr>
<td>vocative</td>
<td>dobry student</td>
<td>dobry koc</td>
<td>dobry stole</td>
<td>dobra szkola</td>
<td>dobre wino</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plural</th>
<th>ci studenci</th>
<th>te koty</th>
<th>te stoly</th>
<th>te szkoły</th>
<th>te wina</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominative</td>
<td>ci studenci</td>
<td>te koty</td>
<td>te stoly</td>
<td>te szkoły</td>
<td>te wina</td>
</tr>
<tr>
<td>accusative</td>
<td>tych studentów</td>
<td>te koty</td>
<td>te stoly</td>
<td>te szkoły</td>
<td>te wina</td>
</tr>
<tr>
<td>genitive</td>
<td>tych studentów</td>
<td>tych kotów</td>
<td>tych stolów</td>
<td>tych szkół</td>
<td>tych win</td>
</tr>
<tr>
<td>dative</td>
<td>tym studentom</td>
<td>tym kotom</td>
<td>tym stolom</td>
<td>tym szkołom</td>
<td>tym winom</td>
</tr>
<tr>
<td>instrumental</td>
<td>tymi studentami</td>
<td>tyimi kotami</td>
<td>tyimi stolami</td>
<td>tymi szkołami</td>
<td>tymi winami</td>
</tr>
<tr>
<td>locative</td>
<td>tych studentach</td>
<td>tych kotech</td>
<td>tych stolach</td>
<td>tych szkolach</td>
<td>tych winach</td>
</tr>
<tr>
<td>vocative</td>
<td>dobry studenci</td>
<td>dobry koty</td>
<td>dobry stoly</td>
<td>dobra szkola</td>
<td>dobre wino</td>
</tr>
</tbody>
</table>

Altogether there are 14 cells, many of which (especially in the plural) show case syncretism. When we examine the number of differences between the agreement classes, we notice that it is minimal between the paradigms of student and stol (3), when compared to, for example,
the number of differences between student and szkoła (10). The minimal number of differences between the first two paradigms indicates that the two classes are subgenders, or one is a subgender of the other. Below we also note how agreement in the predicate indicates the presence of two genders in the plural: -li (masculine personal), and -fy (non-masculine personal). Table 5 shows four separate sets of predicative agreement forms: masculine animate (personal), masculine animate (non-personal), feminine and neuter. If we now add attributive agreement forms for the two numbers in the accusative, we will have five genders in Polish (Table 6). So far, scholars’ opinions as to the number of genders in Polish have not been consistent. The debate, which began more than half a century ago, is ongoing. At the same time, the number of declensional classes in Polish remains a subject open for further discussion.

<table>
<thead>
<tr>
<th>Table 5. Predicative agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>student ‘student’</td>
</tr>
<tr>
<td>Singular</td>
</tr>
<tr>
<td>‘student stood’</td>
</tr>
<tr>
<td>Plural</td>
</tr>
<tr>
<td>-l</td>
</tr>
<tr>
<td>-li</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. Predicative and attributive agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMINATIVE</td>
</tr>
<tr>
<td>Predicative</td>
</tr>
<tr>
<td>Singular</td>
</tr>
<tr>
<td>Plural</td>
</tr>
<tr>
<td>ACCUSATIVE</td>
</tr>
<tr>
<td>Attributive</td>
</tr>
<tr>
<td>Singular</td>
</tr>
<tr>
<td>Plural</td>
</tr>
</tbody>
</table>

1.1.4 Agreement

Agreement in Polish serves as both a basis for grammatical gender assignment and the evidence for the presence of semantic as well as grammatical gender in Polish. Consider the following examples:

(15) Pola/Tadek bought a nice bike. Later she/he sold it.
In English, the verb form remains the same regardless of the gender of the subject noun. The only change can be observed in the second sentence where the anaphoric pronoun has to agree with the referent from the first sentence. There is no need for the agreement between the noun bike and the adjective nice. In contrast:

    Pola(F) buy-PST.F pretty-M bike(M) then him sell-PST.F
    'Pola bought a new bike. Then she sold it'

Tadek kupił ładny rower. Potem go sprzedał.
    Tadek(M) buy-PST.M pretty-M bike(M) then him sell-PST.M
    'Tadek bought a new bike. Then he sold it'

In the Polish example (ex. 16), the verb kupiła 'buy-F.PST' must agree with the feminine name Pola, while kupił 'buy-M.PST' must agree with the masculine name Tadek. Furthermore, the adjective ładny 'nice-M NOM' which pertains to the masculine inanimate noun rower 'bike(M)-ACC' has to follow its gender and take a masculine form as well. As there is no reference within the second clause with which the pronoun go 'he-ACC' and the verbs sprzedała 'sell-F.PST' and sprzedał 'sell-M.PST' can agree, the agreement goes beyond this clause, and allows go to agree with rower, sprzedała with Pola and sprzedał with Tadek.

Agreement in Polish is not as straightforward as may see from the above examples, as will be seen later when hybrid nouns, resolution and nouns involving reference problems are discussed. We shall consider here all elements showing agreement in Polish:

<table>
<thead>
<tr>
<th>Table 7. Attributive agreement in Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
</tr>
<tr>
<td>attributive</td>
</tr>
<tr>
<td>Adjectives</td>
</tr>
<tr>
<td>predicate</td>
</tr>
<tr>
<td>'house is big'</td>
</tr>
<tr>
<td>Pronouns</td>
</tr>
<tr>
<td>personal</td>
</tr>
<tr>
<td>possessive</td>
</tr>
<tr>
<td>'that's my house'</td>
</tr>
<tr>
<td>relative</td>
</tr>
<tr>
<td>'that's the house, which...'</td>
</tr>
<tr>
<td>Participle</td>
</tr>
<tr>
<td>passive</td>
</tr>
<tr>
<td>active</td>
</tr>
<tr>
<td>'standing house'</td>
</tr>
</tbody>
</table>

38
The agreement limitations in syntax have been briefly discussed in §1.0.2, and an example from German has been given, where adjectives agree with their head noun, but not in the predicate position. This is not the case in Polish. All adjectives take the gender of the nouns in the attributive as well as predicate position; hence, there is no limitation in syntax in agreement between nouns and adjectives. The fact that Polish verbs take gender agreement in the past tense only is regarded as an interaction between tense and agreement.

<table>
<thead>
<tr>
<th>Numerals</th>
<th>Pierwszy dom</th>
<th>Pierwsza lampa</th>
<th>Pierwsze łóżko</th>
</tr>
</thead>
<tbody>
<tr>
<td>'first house'</td>
<td>'first lamp'</td>
<td>'first bed'</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demonstratives</th>
<th>Ten dom</th>
<th>Ta lampa</th>
<th>To łóżko</th>
</tr>
</thead>
<tbody>
<tr>
<td>'this house'</td>
<td>'this lamp'</td>
<td>'this bed'</td>
<td></td>
</tr>
</tbody>
</table>

Before we proceed to more challenging cases, we shall once more refer to the Agreement Hierarchy described by Corbett (1991: 226), which has been based on the assumption that “as we move rightwards, the likelihood of semantic agreement will increase monotonically (with no intervening decrease)”: 

attributive < predicate < relative pronoun < personal pronoun

In order to see this process, in looking at examples below “we shall begin with cases where syntactic agreement is dominant, and then progress to those where semantic agreement has a greater role” (Corbett, 1991: 226). The first type of problematic cases are hybrid nouns, i.e. nouns taking more than one kind of agreement. Whenever hybrid nouns occur, a special role is played by pronouns, which clarify their referents by taking an appropriate form (usually ona ‘she or on ‘he’). Semantic agreement which these nouns take is always consistent with the gender assigned by the semantic rules. In the example below, the second clause contains a “dropped” pronoun that corresponds with the semantic gender of doktor. In such cases, the speaker must possess some necessary knowledge about the sex of the doctor to be able to use the correct gender on the target (here the adjective zajęta ‘busy’).

(17) Doktor-ą Nowak jest w gabine-cie
docent(F)-NOM Nowak be.PRS.3SG in office(M)-LOC.SG
Jest dość zajęt-ą
be.PRS.3SG quite busy-F
'Doctor Nowak is in the office. She is quite busy.'
Suppose the receptionist gives information about Doktor Nowak, whom the patient has never met before. Since the surname does not provide any indication as to the sex of the referent, the patient is faced with a choice of a masculine and a feminine form of the adjective:

RECEPTIONIST:

(18) Doktor-o Nowak jest w gabine-cie.
doc-tor(F)-NOM Nowak be-PRS.3SG in office(M)-LOC.SG
'Doctor Nowak is in the office.'

PATIENT:

(19) Czy jest bardzo zajçt-a/y?
Q be-PRS.3SG very busy-F/M
'Is she/he very busy?'

Other instances of hybrid nouns can be exemplified by the noun *dziewczę* 'girl(N)' and *lampa* 'fumbler'. The former arises from the conflict of the assignment rules, similarly to the often-quoted example from German *das Mädchen*. According to Corbett (Corbett, 1991: 227), "Many Indo-European languages assign sex-differentiable nouns to the masculine or feminine gender as appropriate, while the young of sex-differentiables - typically young animals which are treated as not yet sex-differentiable are neuter". The noun *dziewczę* denotes a woman and should be feminine, but it also denotes a young being, and is thus neuter. The neuter assignment is also further supported with the fact that the noun *dziewczę* ends with the vowel *ę*, which is typically a neuter ending in Polish (other neuter nouns like *ciełu* 'calf(N)', *zielę* 'herb(N)'). The complex situation is reflected in the agreement, which the noun can take in sentences below.

(20) Widzial-es t-o sliczn-e dziewczę, ja k
see.PST-M.2SG that-N beautiful-N young, girl how
ładnie spiewa-l-o?
nicely sing-PST.IPFV-N.3SG
'Did you see that beautiful young girl, how nicely she was singing?'

The demonstrative *to* 'that', the attributive modifier *sliczne* 'beautiful', and the predicate *spiewało* 'sing.PST.IPFV-N.3SG' all indicate the neuter agreement. Interestingly, contrary to German *das Mädchen*, the personal pronoun in the example below does not allow a choice between neuter and feminine agreement (the semantic agreement is obligatory here).

(21) Jak to dziewczę się dużo uczy!
how this young, girl(N) REF a lot learn.PRS.3SG
Ona kiedyś zostanie profesor-em.
she one.day become.FUT.3SG professor(M)-INS
'This girl learns a lot! She will become a professor one day.'
Of great interest is the masculine form *profesor* used with reference to *dziewczę*, which we have just seen being assigned neuter gender on the basis of the formal and semantic criteria. Nouns exemplified by *profesor* and *doktor* belong to the group of epicene nouns and can denote both masculine and feminine referents. Theoretically, thanks to the derivational means available in Polish, a noun *profesorka* could be used as a feminine equivalent of the noun *profesor*. It must be noted, however, that for some occupations, this kind of transformation of the job title from masculine to feminine gender attaches derogatory meaning to the newly created noun.

We will now briefly analyse a few derogatory terms which are meant to refer to men, but which have the morphology of nonmasculine personal or both masculine personal and non-masculine personal gender. The often-discussed case of *lajdaki* 'rascal' serves as an example of the problem, but there are many more similar nouns (*bandzior* 'bandits', *drybłasy* 'strappers', *typy* 'types', *pryki* 'old farts', *obiboki* 'layabouts', *obdartusy* 'ragamuffins', *pikiusy* 'guzzlers', *nieroby* 'skulks', *lenie* 'idlers' etc.). *Lajdaki* takes nonmasculine personal agreement in attributive and predicate position (Corbett, 1991: 234):

(22) Te lajdaki znów mnie oszuka-ly!

Those rascals cheated on me again!

Adjectives often take agreement *ad formam*, i.e. they take non-masculine personal agreement, while pronouns take agreement *as sensum*, taking masculine personal agreement (Rothstein, 1976: 248):

(23) Ci co się poda-ją za bezrobojnych

Those who report themselves as unemployed

to są zwykle lenie śmierdzące

that be typical idlers stinking

którzy nie chcą się zajmować prac

who don't want to be stakeholders

‘Those who report themselves as unemployed are typical stinking idlers who don’t want to deal with a job.’

It is possible to notice how the Agreement Hierarchy is being applied in the cases discussed above. As we move rightwards, the likelihood of semantic agreement indeed increases. Let us consider one last example provided by Rothstein from a story by Antoni Słonimski, who describes how he was teaching Polish to an artist:
The beast was so talented that by the morning he spoke (the language) too well.'

Rothstein (Rothstein, 1976: 248) comments on this case in the following way:

Przykład ze Słonińskiego ilustruje...pewną tendencję składniową, polegającą na tym, że w obrębie zdania pojedynczego zwycięża zwykle rodzaj gramatyczny, lecz im bardziej oddalone od danego rzeczownika...będą wyrazy wiązające się z nim syntaktycznie, tym prawdopodobniejsza będzie przewaga rodzaju naturalnego'.

The emphasis in bold in ex. 24 has been applied to Rothstein’s intriguing example to indicate the grammatical agreement between the head noun bestia ‘beast’ and the verb in the past była ‘was’, both of feminine gender. The two elements stand in syntactic proximity defined by Rothstein, as opposed to bestia and the verb mówił ‘spoke-M’, which are placed at some distance from one another; hence, the semantic agreement takes precedence over the grammatical agreement.

Mixed semantic and syntactic gender resolution in Polish is another source of troublesome agreement. According to the rules for predicate agreement in Polish, in the plural, non-masculine personal nouns take agreement forms in -y (szły ‘went-N_M.PERS’), while masculine personal nouns take forms in -i (szli ‘went-M.PERS’). Conjoined NPs provide evidence that if one (or more) of them is a masculine personal noun, the masculine personal form in -i is used, and in all other cases, i.e. if none of the NPs is headed by a masculine personal noun, the -y form is used. In general, conjoined NPs in Polish act accordingly, with some exceptions. Consider this example from Doroszewski (1962):

(25) Hania i Reks bawi-li siê pilk-q
Hania(F) and Reks(M.ANIM) play-PST.IMPF.M.PERS REF ball(F)-INS.SG

‘Hania and Reks played with a ball.’

The bolded suffix -li is evidence that the masculine personal form has been used, despite the fact that neither Hania nor Reks are masculine personal. Similarly, the above

19 An example from Słoniński illustrates...a syntactic tendency, according to which in a simple sentence grammatical gender usually wins, but the further from the noun the words connected to it are going to be, the more probable that the semantic gender will win [translation mine].
stated rules can be violated when a feminine form denoting a human is conjoined with a masculine inanimate noun (Corbett, 1991):

\[26\] **Mama, córeczka i wózek**

mother(F) daughter(F) and pram(M.INAN)

*ukaza-li się nagle*

show-PST.IMPF.M.PERS RBF suddenly

'The mother, the daughter and the pram showed suddenly.'

Clearly, new rules of agreement are necessary to cover the above-described exceptions. Corbett (1991: 286) suggests the following set:

I. If the subject includes a masculine personal conjunct, the predicate will be in the masculine personal form

II. (optional) If the subject includes the features of masculine and personal, whether these are syntactic or semantic, the predicate may be in the masculine personal form

III. (optional) If the subject includes a masculine animate conjunct, the predicate may be in the masculine personal form

IV. Otherwise, the predicate will be in the non-masculine personal form

Interestingly, according to Corbett (1991: 296), "the use of masculine personal as the semantically justified gender can be explained on the same semantic grounds as for French and Slovene". The title *panstwo* or the collective phrase *panie i panowie* "ladies and gentleman" both take masculine personal forms. Moreover, Polish, similarly to English, also uses generic masculine forms, which may serve as an additional justification for favouring masculine agreement in the resolution examples.

Polish uses generic masculine forms with nouns involving reference problems. Gender agreement in these cases can be puzzling, as it emerges from a simple example:

\[27\] **Pacjent powinien wiedzieć, że jego obowiązkiem jest...**

patient(M)-NOM.SG should know.INF that his duty(M)-INS.SG be.PRS.3SG

'A patient should know that his duty is to...'

Examples such as (27) above, where the form *jego* is used generically with reference to both sexes, are very common in the Polish language. Such use results from purely formal factors: *pacjent* is masculine, and thus the pronoun is masculine as well. Interestingly, despite that fact that there exists an equally common feminine form *pacjentka*, it is unlikely to occur in sentences such as:

\[^{20}\text{Based on Zieniukowa, J. (1979). Składnia zgody w zdaniach z podmiotem szeregowym we współczesnej polszczyźnie. *Slavia Occidentalis* (36), 117-129.}\]
However, a collective way of referring to both sexes may appear, although it is also based on masculine personal forms:

(29) Pacjenci i pacjentki powinni wiedzieć, że ich obowiązk-iem jest...

'The (male) patients and (female) patients should know that their duty is to...

Agreement in Polish supports Corbett's claim that there is no one point at which agreement can be neatly divided (Corbett, 2001: 8-9), mostly due to the fact that so many different elements show agreement in so many clear and unclear circumstances. Thus any discussions concerning the boundaries of agreement are bound to conclude the same. To sum up, the preceding paragraphs have been aimed at collating as many aspects related to gender in Polish as possible, without blurring the picture, but rather, with the purpose of comparing different theories and approaches to those aspects, and showing the richness of this category.

1.2 Gender in English

The debate on agreement boundaries, which has been mentioned in the preceding section, is directly relevant to the category of gender in English. Some linguists (e.g., Wiese, 1983) express a view that pronominal anaphora do not belong to the local domain of agreement and should therefore be excluded from agreement boundaries. If this view were generally accepted, English would be regarded as a genderless language, since the only way English can manifest the category of gender is via personal pronouns. Corbett (1991: 169) gives two arguments supporting the idea of a gender system in English. Firstly, “when language marks gender on pronouns and on some other target type, then typically they require a similar machinery to handle them all” and secondly, “when pronouns are the only evidence for gender, then the resulting gender system seems to be of the same type as that found in some other, fuller agreement system.” Although English gender is restricted to pronouns, there are a few issues worth considering.
1.2.1 Gender assignment

In English, in contrast to Polish, the lack of declension, reduced noun inflection and verb conjugation make the inflectional system particularly simple, since many inflectional categories either lack overt marking, or are expressed periphrastically. As to gender assignment, in standard English spoken in the UK there are no formal criteria, but only strict semantic assignment rules: masculine gender is assigned to male humans, feminine gender is assigned to female humans, and neuter gender includes the residue. There are exceptions that involve upgrading of inanimates to masculine or feminine gender or downgrading of animates to neuter gender.

Several studies\(^{21}\) have investigated non-standard varieties of English which show gender assignment rules based on a mass-count distinction on nouns, where count nouns can be referred to anaphorically with “gendered pronouns” (i.e. he, him; she, her), whereas mass nouns only employ neuter it. Such “gendered pronouns” are found in two varieties in particular, namely traditional “West Country” (Southwest) dialects and (West-Country based) Newfoundland English. In these varieties, neuter pronouns are traditionally only employed for mass nouns, while count nouns trigger masculine forms. In modern spoken English, however, the “gendered pronoun” of choice is feminine rather than masculine. As a result, linguists investigating gender assignment in English today are faced with three interacting and sometimes conflicting systems:

1) The Standard English system, at least in writing, demands neuter pronouns for inanimate nouns, while humans (or animate entities in general) are either he or she according to sex. There are only few exceptions to this rule. For example, ships and some other vehicles can be feminine metaphorically.

2) In the traditional West Country system, based on a semantic distinction in the referent nouns (mass vs. count), mass nouns trigger neuter forms, while count nouns use masculine pronouns. Feminine forms only occur with humans (women).

3) In basically all varieties of spoken English, speakers use feminine forms in particular when they want to imply a certain degree of emotional involvement (positive or negative) with what is being said. This use seems to extend not only to concrete, but also to abstract references, such as situations or circumstances.

Despite its simplicity, the gender assignment system in English can still be intriguing. There are a few special referent classes, such as personification (upgrading), where the referent is assimilated to a human being (is being upgraded), and reference to animals, which require more detailed analysis. Wagner (2002) deals with these two notions in much greater detail than is required for the purpose of this thesis, so we shall only refer to

\(^{21}\) Information on varieties of English based on Wagner (2002).
some of her examples and discussion. According to Wagner's comprehensive study on English pronouns (2002: 120), the examples of personification in English are restricted to myths, legends and children's stories. She excludes from personification instances such as using he by the cider maker when referring to an apple, or by the watchmaker when referring to one of the watches. Wagner claims that personification has usually been associated with feminine forms, while in instances just mentioned, the pronoun he is used in the vast majority. A well-known example of personification in English is the use of the pronoun she with reference to a ship, which is "probably based on the imagery of a ship as a womb-like container" (Wagner, 2002: 121).

Mathiot (1979), on the other hand, argues that the instances of upgrading recorded in her database are extremely numerous. She adds that "any nonhuman entity can be referred to as either 'he' or 'she'," and enumerates the items which can be upgraded: ice cream cone, a football team, mathematical formulae, high prices, a California poppy, grass, a pillow, a vase, a key, a door, etc. Animal referents constitute another rich source of the expansion of the gender system in English. Speakers of English use personal pronouns he and she with reference to their pets, and according to Wagner (2002: 121), "a more detailed investigation...reveals that the observed pattern is the rule rather than exception." Pronoun switches seem to be something typical in English, especially when the domestic animals are named and have a special place in family life. Emotive or affective use of pronouns is an extremely rich source of information about how speakers of English use gender of personal pronouns with reference to animals, objects and people. Mathiot (1979) provides the following examples:

(30) Can you keep her there until I can pick her up?
    (the speaker is a woman ordering a slipcover on the telephone; the suspected meaning of she was 'prized possession')

(31) Oh, he's still good for a few more years!
    (the referent is a battered raincoat; the suspected meaning of he is either 'competent' or 'good-natured')

(32) A woman complained that her uncle referred to her small grey cat as he despite the many times she had told him the cat was a female (the suspected meaning of he is either 'ugly' or 'good natured')

Wagner (2002: 124) highlights certain aspects of pronominal use connected with animal referents. In the following example, a police officer is being questioned about dogs on the force. The formal character of the interview coupled with the fact that the police officer has never owned such a dog, explain the use of the pronoun it while referring to a dog. It changes into him once the police officer becomes more emotionally involved and begins to talk about a dog becoming a family member:

46
Alright? Next question. Yes young man.
[PS0000]: What was it like when you had your police dog?
[PS11SF]: I have never had a police dog. I've never had, never been on the special force. A lot of people like it... because basically when you look after a police dog it becomes your pet as well, you take it home with you, and you take it to work with you, and you’ll have a police dog for sort of like its working life is of seven to eight years, so basically you’re gonna have him for seven to eight years and he becomes like a family pet.

The choice of a pronoun may also be affected by pragmatic factors, e.g. politeness (we may decide to use a personal pronoun he or she with reference to our friend’s pet). Wagner (2002) explains that even being a “cat person” or a “dog person” may affect people’s choice as to which pronoun to use, i.e. “cat people” are likely to refer to the dog that chased their cat as it rather than as he or she, and vice versa.

Downgrading (opposite to upgrading) corresponds to the denial of human status. It indicates various degrees of negative involvement, which may range from “lack of interest to mild annoyance, contempt, or even violent rage” (Mathiot, 1979). Mathiot also gives a “striking example” of downgrading of an item, which is normally upgraded:

(33) Sometimes I feel like junking it, just tossing... But then she comes back with her choke working okay. I just don’t know what I am going to do without her.

(33) Sometimes I feel like junking it, just tossing... But then she comes back with her choke working okay. I just don’t know what I am going to do without her.

(33) Sometimes I feel like junking it, just tossing... But then she comes back with her choke working okay. I just don’t know what I am going to do without her.

(33) Sometimes I feel like junking it, just tossing... But then she comes back with her choke working okay. I just don’t know what I am going to do without her.

1.2.2 Gender agreement

Gender agreement in English is limited to: semantic agreement between head nouns and anaphoric/cataphoric pronouns, hybrid nouns, and generic use of the pronoun he. One of the sources of problems with gender agreement in English is the fact that most nouns do not distinguish between male and female referents. Therefore, the question of how to overcome the difficulty of the unknown referent appears whenever a person says:

(34) I’m going to see my new doctor today. I hope he/she is nice.

(34) I’m going to see my new doctor today. I hope he/she is nice.

(34) I’m going to see my new doctor today. I hope he/she is nice.

(34) I’m going to see my new doctor today. I hope he/she is nice.

There are instances of replacing the bulky “he or she” phrase with the 3rd person plural they as a neutral pronoun. Furthermore, in comparison to Polish, which is regarded as the most sex-biased of Slavonic languages, due to the masculine personal agreements taking over in case of any conflict agreement situation, English is regarded as not sexist at all. However, consider ex. (35):

(35) Everyone loves his mother.

(35) Everyone loves his mother.

(35) Everyone loves his mother.

(35) Everyone loves his mother.
Corbett explains that the use of *he* in cases like this is an example of the “generic use of the masculine pronoun to denote both males and females” (1991: 221). Nouns such as *chairman* or *congressman* have been replaced with non-sex-differentiable equivalents, such as *chairperson*. Agreement between noun phrases involving reference problems and their anaphoric pronouns in English largely depends on the speaker's knowledge of the agreement target. Pronouns in English are central to the gender system, and they are also the basis for dividing English nouns into different agreement classes.

Although she at first claims that personifications in English are restricted to myths, legends and children’s stories, Wagner admits that native speakers of English of all the investigated varieties are not happy about the ubiquitous status that *it* has assumed in their mother tongue. Wagner's educated guess is that the speakers choose to employ personal pronouns other than semantically empty *it* if they want to add “feeling” to an utterance, regardless of whether the actual emotion is positive or negative. Clearly, such motivation goes beyond myths and legends, and demonstrates that the use of pronouns by English native speakers is far more creative and semantically motivated than one may have suspected. Wagner concludes that the “natural gender system of StE [Standard English] thus seems to be on its way to becoming a “pragmatic gender system” in the Spoken Standard, where forms marked for gender are used according to the different requirements, emotionality and general circumstances of the situation.

1.3 Conclusions

This chapter was aimed at providing an overview of aspects related to gender systems in Polish and English. The two systems have only one thing in common: semantic gender assignment, which is one of the aspects analysed in this thesis. Generic uses of pronouns, hybrid nouns, upgrading (personification), downgrading and other issues discussed in this chapter are also a very good platform for a comparative gender analysis when two languages are learned side by side. However, these areas are not discussed in this thesis. We have considered the role of gender in English and Polish and we will now look at how gender is acquired.
Chapter Two. Acquisition of gender

2.0 Introduction

A study of the acquisition of grammatical gender by children requires a closer look at the work that has been done in the area of child language development to date. The aim of this chapter is to discuss theories and research on child language acquisition, and as a result, show how studies on the acquisition of grammatical gender in mono- and bilingual children have evolved. This chapter is divided into two parts: the first section discusses language acquisition and the theories that played the most important role in the formation of the present research in the field. The second part examines the studies on the acquisition and development of grammatical gender in mono- and bilingual children.

2.1 Language development

Some Greek philosophers (Plato, Herodotus, Augustine) were intrigued by the mystery of how babies learn language, but the research on language acquisition as "an empirical enterprise" only dates back to the second half of XIX century, when diary studies were the main way of collecting data. Early research reveals that scholars showed a considerable interest in speech comprehension, which receded as the focus of research shifted to language production (Fernald, 2002: 203). The first diary was published in 1787, and more followed. They provided anecdotal data on linguistic development, but there were also rich in observations, which were of scientific interest. Because of their great interest in the children’s understanding of language rather than the production of actual words, these early scholars observed a clear discrepancy between comprehension and production in the children’s speech. They provided rich data, "contributed insights about first-language learning that were often overlooked in the more carefully controlled research that followed" and "their commentaries ranged widely, unconstrained by received opinion about which
questions were worthy of investigation" (Fernald, 2002: 105). The originality of those first child language studies cannot be disputed, despite the fact that current research standards and scientific rigour, which emphasize the systematic data collection and verifiable data analyses, are undoubtedly much higher than they were in the past centuries. After the diary period of language studies, scholars began analysing speech using large samples, which were cross-sectional and much more systematic than diary studies. Studying large number of subjects (as many as 480 children) allowed collecting only small samples per child, which permitted quantitative rather than qualitative analysis. Around the mid 1960s, the method of research in child language development area shifted from large sample studies to longitudinal and experimental studies, which usually included three children studied regularly, and resulted in a growing number of case studies. The children involved were audio-recorded, and the transcribed data was handwritten. Since the beginning of the era of personal computers, data has been transcribed onto computer files. Major studies associated with this new observational trend were Braine (1963), Bellugi (1967), Bloom (1970) and Brown (1973). Longitudinal and experimental methods remain the most popular and the most reliable way of gathering both linguistic and non-linguistic data from children. The majority of studies that have been carried out since those studies mentioned above are either just longitudinal (case-studies), just experimental, or both. The combination of the two methods is regarded as the way to obtain the fullest account of child speech data. Years of longitudinal research and transcription led to the emergence of The Child Language Exchange System (CHILDES) created by MacWhinney and Snow (1985)\(^2\).

The interpretation of child language data with respect to any aspect of language largely depends on our approach to theories and perspectives existing in the area of language acquisition and development. Throughout the history of child language research, there have been numerous theoretical approaches and perspectives, which have guided the data analysis and the interpretation of the results. The early diary studies did not follow any particular theoretical orientation, but scholars dealing with large sample studies were influenced by the first proposal in the modern context, which was the theoretical orientation called *behaviourism*. Behaviourists believed that children are born as “blank slates”, i.e., babies (of all species) are born without any knowledge whatsoever, and they “learn pieces of language by means of instrumental conditioning (based on principles of association) and...they generalize to new instances by means of stimulus generalization (based on principles of induction)” (Tomasello, 2003: 2). Behaviourist psychologists developed their theories while carrying out a series of experiments on animals (e.g., rats or birds), which, as they observed, could be taught to perform various tasks by encouraging habit-formation. Researchers

\(^2\) More information on CHILDES to follow in Chapter Three.
rewarded desirable behaviour, which was known as positive reinforcement. Undesirable behaviour was punished or not rewarded at all, and was known as negative reinforcement. The main representative of behaviourism and its influence on language studies was Skinner (1957), whose ideas were reflected in terms such as “acquisition of vocal habits”, nowadays simply called language acquisition. According to Skinner, a child imitates parents and carers. Adults either recognise words spoken by the child, and reward the child’s successful attempts (positive reinforcement), or forget the unsuccessful trials (negative reinforcement). Skinner also suggested that the successful use of a sign such as a word or lexical unit, given a certain stimulus, reinforces its “momentary” or contextual probability. This learning perspective presupposed that there was no need to refer to such hypothetical constructs as the mind in order to explain language acquisition, because children learn language as a response to stimuli, just as rats learn to tackle the maze in the same way. In his well-known book *Verbal behaviour*, Skinner (1957: 3) stated:

> The basic processes and relations which give verbal behaviour its special characteristics are now fairly well understood. Much of the experimental work responsible for this advance has been carried out on other species, but the results have proved to be surprisingly free of species restrictions. Recent work has shown that the methods can be extended to human behaviour without serious modifications.

Skinner’s conclusions did not seem to be shared by many, according to Chomsky (Skinner’s biggest critic), who pointed out that, at the time, in many publications of confirmed behaviourists there was “a prevailing note of scepticism with regard to the scope of these achievements” (Chomsky, 1959). While there may be some truth in Skinner’s explanation, there are many objections to it too. The vast majority of children go through the same stages of language acquisition. There appears to be a definite sequence of steps, which psycholinguists refer to as developmental milestones. The sequence seems to be largely unaffected by the treatment the child receives or the type of society in which she grows up. Moreover, there is evidence for a critical period for language acquisition. Children who have not acquired language by the age of about seven will never entirely catch up (cf. feral children). Another problem with behaviourism in terms of language acquisition is the fact that children are often unable to repeat what adults say, especially if their utterances contain structures which the child has not yet started to use. The classic demonstration comes from McNeill (1973). The structure in question here involves negating verbs:

C: Nobody don't like me
M: No, say, "Nobody likes me."
C: Nobody don't like me.
(More repetitions of this dialogue)
M: No, now listen carefully: say, "Nobody likes me."
C: Oh! Nobody don't likes me.
Additionally, few children receive much explicit grammatical correction, since parents are more interested in politeness and truthfulness. According to Brown at al. (1969: 71),

It seems to be truth value rather than well-formed syntax that chiefly governs explicit verbal reinforcement by parents — which renders mildly paradoxical the fact that the usual product of such a training schedule is an adult whose speech is highly grammatical but not notably truthful.

The fact that nearly all children arrive at correct answers even though their mistakes are not corrected shows that the argument of the children learning from verbal stimuli is not a complete theory. This shows that language is based on a set of structures or rules which could not be worked out simply by imitating individual utterances. The mistakes made by children reveal that they are not simply imitating but actively working out and applying rules. For example, a child who says "goed" instead of "went" is not copying an adult, because such a form would surely not appear in the input. She is rather over-generalising a rule, which she has discovered: the past tense verbs are formed by adding a /d/ or /t/ to the base form, and the "mistakes" occur because there are irregular verbs which do not behave in this way. Such forms are often referred to as intelligent mistakes.

Behaviourism as a research programme was divided into several types, but of one sort or another, it was an extremely popular "methodological commitment among students from about the second decade of the twentieth century through its middle decade, at least until the beginnings of the cognitive science revolution" (Graham, 2007), which is believed to have been a direct result of Chomsky's scathing review of Skinner's *Verbal Behavior* (Chomsky, 1959). With reference to language acquisition, in addition to some of the earlier arguments, Chomsky believed that children could not possibly learn a language by means of simple association and induction, due to the abstractness and often arbitrariness of some principles of grammar. He stated that the child hears only a sample of language from her parents, who do not always speak in grammatically complete sentences, and hence the child receives impoverished language input (the so-called poverty of the stimulus argument), which does not provide all the necessary structures for a child to be able to learn them from adults. Thus, Chomsky concluded that children must have an inborn faculty for language acquisition. According to his theory, known as the innateness theory, as a natural predisposition, hearing speech triggers the language acquisition process in children, and the child’s brain is able to interpret what she hears according to the underlying principles or structures it already contains. This process is biologically determined, i.e. at birth, children’s brains are equipped with a special language mechanism. In other words, our brains are prewired to learn languages. This natural faculty has become known as the Language Acquisition Device (LAD). Chomsky’s claim should not be interpreted as meaning that, for
example, a Polish child is born knowing anything specific about Polish. He stated that all human languages share common principles, e.g., they all have words for actions (nouns or verbs), and the child has a task to establish how these underlying principles are expressed in the language she hears.

The argument of the poverty of the stimulus and the resulting hypothesis that "human beings are born with an innate universal grammar containing a number of abstract principles that guide the acquisition process" (Tomasello, 2003: 2) had a profound effect on child language researchers in 1960s and 1970s. The atmosphere of the changes that were taking place in those years is described by Tomasello (2003: 2):

The prevailing opinion at the time was that baby utterances such as "More juice" and "Doggie gone" were just that, baby utterances that rested on very concrete and seemingly non-adult-like linguistic representations such as More X and X gone (e.g. Braine, 1963). But people impressed with the argument from the poverty of the stimulus looked at those baby representations and at the formal descriptions of adult language being proposed by Chomsky and others and said, in effect: "You can't get there from here" (e.g., Gleitman and Wanner, 1982). The majority of opinion in the field thus changed rather quickly to the view that children's early language was somehow undergirded by some kind of linguistic abstractions - perhaps even the same ones that underlie mature adult language. There is the so-called **continuity assumption**: that basic linguistic representations are the same throughout all stages of child language development - since they come ultimately from a single universal grammar (Pinker, 1984).

Chomsky's idea of an innate language faculty found support in several areas of language study. Neuroscientists have identified specific areas of the brain which deal with specifically linguistic functions, i.e. Broka's area and Wernicke's area. Also, research in sign languages show that those are complex and fully grammatical languages, and more importantly, children learning a sign language pass through similar stages to hearing children learning spoken languages (Newport and Meier, 1985, Emmorey, 2002). Those and other studies indicate that there is an innate ability in humans to acquire languages.

According to Chomsky's original position, LAD contained specific knowledge about language. However, despite the fact that other linguists accepted Chomsky's idea of universal grammar, they suggested that children are not born with a set of linguistic categories, but rather with a mechanism to process the linguistic information. Theories that oppose Chomsky's Universal Grammar (UG) suggest that there is less innate structure and more actual learning of the language. Many such theories have emerged in the past 40 years, but there is still no consensus as to the way humans acquire languages as children. Although current approaches are more focused on the child, there is still a clear division between **nativists** (Chomsky and his followers), who propose that children have a pre-wired capacity separate from other mental abilities, and **empiricists** (Plunkett, Tomasello, Slobin),
who accept the existence of the "acquisition device", but do not assume any inborn knowledge, deep structure grammar or abstract linguistic features universal to all languages. Scepticism concerning UG has been present since its origin, and various researchers have modified Chomsky’s theory, openly criticized or simply rejected it. The criticism has usually been based on the fact that Chomsky's work on language is theoretical. His aim is building a theory of language, and since he is interested in grammar and much of his work consists of complex explanations of grammatical rules, he has never, in fact, studied real children. His idea of the way children acquire various features of language ignores the interaction between the child and the carers, relying only on the child's exposure to the language.

Since Chomsky's criticism of Skinner's behavioural theories in the sixties, language acquisition researchers have been choosing between nativist and empiricist approaches to development. In the seventies, "cognitive and social knowledge was introduced as a support system to the language acquisition process in order to take the strain from the overloaded formal learning machine" (Plunkett, 1995: 36). Piaget’s investigations and theories, which placed acquisition of language within the context of a child's mental or cognitive development, formed the basis of the cognitive approach to language development. Piaget argued that a child needs to understand a concept before she can acquire the particular language form which expresses that concept (Piaget, 1972). According to the cognitive theory, infant's cognitive development should be regarded in terms of the growth of representation or the "object concept." As an illustration, let us consider a child who is able to compare objects with respect to size (seriation). Piaget claimed that if a child has not yet reached this stage, she would not be able to learn to use comparative adjectives like "longer" or "shorter". Another phenomenon (object permanence) often referred to in relation to the cognitive approach comes from one of Piaget's stages of development. Before turning one, children seem unaware of the fact that objects exist independently of their perception, i.e. even though they cannot see them. However, around the age of 18 months children show the ability to understand the fact that objects, which move out of their sight, do not simply cease to exist. The cognitive approach points to the noticeable increase in children's vocabulary at that age, and links the fact that children start using more labels for objects at exactly the same time when they begin to understand the idea of object permanence.

The cognitive approach, as well as the subsequent, data-driven, theories (interactionism, usage-based, connectionism, constructivism) have placed great emphasis on the ways in which real children develop language to fulfil their needs and interact with their environment, including other people. Undoubtedly, Chomsky's revolutionary approach remains at the centre of the debate about language acquisition. However, researchers who reject it are authors of new theories, which are child-directed and child-focused. This
includes, among many others, work by Bowerman (1990), Tomasello (1995), Lieven (1997), Bates (2001). Contrary to what Chomsky's followers suggested, “children can get from here to there, and they do it without the aid of any hypothesized universal grammar” (Tomasello, 2003: 3). In contrast to the work of Chomsky, recent scholars have drawn their attention to the importance of the language input children receive from their carers. The point of learning a language is to be able to communicate with other people, which is undoubtedly one of the first functions of language that children are able to observe. It can only be learned in the context of interaction with other people. Interactionists suggest that the language behaviour of adults when talking to children (child-directed speech, CDS) is specially adapted to support the acquisition process, and is often described as scaffolding for the child's language learning. In response to Chomsky's LAD, interactionists coined the term Language Acquisition Support System (LASS).

Some empiricists have argued that language can be acquired using general-purpose perceptual learning rules (such as pattern recognition, association, and conditioning). One of the most recent views of language development and human linguistic competence is the usage-based approach, also referred to as cognitive-functional linguistics (e.g., Langacker, 1987; Croft, 1991; Langacker, 1991; Bybee, 1995; Tomasello, 1998). It is a group of theories which view language as a “window on the operation of the human mind”, in which “the patterns of language emerge not from a unique instinct but from the operation of general processes of evolution and cognition, social processes, and facts about the human body” (MacWhinney, 2005: 257). In other words, language structures emerge from language use. As opposed to UG and other formal approaches, usage-based theories hold that “the grammatical dimension of language is a product of a set of historical and ontogenetic processes referred to collectively as grammaticalization” (Tomasello, 2003: 5). Hence, grammatical rules are not just “algebraic procedures for combining words and morphemes that do not themselves contribute to meaning” (Tomasello, 2003: 5), but they are in fact meaningful linguistic symbols, which construct language.

2.2 Bilingual language acquisition

Research in bilingual children may not have had a history as long as research in monolingual children, but the results obtained from the field are regarded by many linguists and psycholinguists to be essential. Various theories of language development have been based on research in monolingual or bilingual cognitive, social, psychological and linguistic development of language. However, the insights from the research in bilingualism have

---

23 Emphasis by Tomasello.
proven crucial for a better understanding of language acquisition processes. Early bilingual studies had the form of diary studies, such as Ronjat (1913), similar to monolingual research. The number of studies dealing with childhood bilingualism increased throughout the 1960s and 1970s, after the publication of the results of a comparative study of bilingual and monolingual children conducted by Peal and Lambert (1962). By 1982 research into childhood bilingualism flourished to such an extent that it was necessary to draw together all the data and organise it into some framework. The aims were to summarise the work on bilingualism and make it accessible to developmental psychologists, and to provide researchers with an integrated model of the developmental processes operating in the bilingual child (Homel, 1987: 3f). Researchers organised their work into several topic areas, such as: language acquisition and processing, cognitive functioning and development, social and emotional development, bidialectism and biculturalism. The classification helped linguists find their way in the main fields of research in bilingualism and resulted in further development of these areas. In recent years, a great number of researchers working with bilingual children have contributed to the field by providing various experimental and longitudinal data. At the turn of the 1980s and the 1990s, the study of the cognitive and linguistic achievements of bilingual children became a popular area of inquiry. The issue was by no means new, but its richness was only to be discovered, especially in such fields as psychology and education. In the introduction to a collection of papers, Bialystok (1991) explains why empirical data of the consequences of bilingualism for children’s cognitive development, linguistic processing and metalinguistic abilities were rare. She claims that one factor to blame for making these aspects so difficult to study is the enormous diversity that accompanies children’s bilingualism. Bialystok emphasizes two changes that resulted in the modification of the research in the linguistic and cognitive development of bilingual children. First is the fact of including second-language studies into the mainstream linguistic and psycholinguistic research, i.e. second-language studies are no longer treated merely as a variation on a theme in language acquisition research. Second is the shift in emphasis in psycholinguistic research from descriptions of the products to the analysis of the processes. In other words, the investigation is based mostly on the analysis of the developmental process instead of only final accomplishments (Bialystok, 1991: 5).

The late 1980s and early 1990s witnessed the existence of two separate approaches as to whether children produce language specific structures at all stages of their development (separate development hypothesis), or mix the two languages into one system at an early stage, and separate them as they become more proficient in both languages (unified system hypothesis). The latter theory was supported in the works by Voterra and Taeschner (1978), who suggested the existence of three stages: in the first stage, the child has one lexical
system, which includes words from both languages, and his linguistic development seems to
be like the one of a monolingual child. In the second stage, the child distinguishes two
different lexicons, but applies the same syntactic rules to both languages. In the third stage,
the child speaks two languages differentiated in both lexicon and syntax (differentiation
happens during the third year of life). Towards the end of the 1990s, the separate
development hypothesis received more support. Researchers believed that if children were
initially monolingual, then they should use their two languages indiscriminately regardless of
context. Evidence shows, however, that bilingual children use their languages in context
sensitive ways (Genesee, Nicoladis & Paradis, 1995, Comeau and Genesee, in press). The
two approaches have been a subject of a heated debate through the last decade of the 20th
century, but with each bilingual study, there is less and less doubt regarding the separate
treatment of the two language systems by bilingual children.

Meisel (1989) was the first to propose a framework known as Bilingual First
Language Acquisition (BFLA). BFLA studies encompass all linguistic research where the
subjects under investigation are acquiring or have acquired two languages from birth. The
defining concept for BFLA is that “a child is exposed to language B no later than a week
after he or she was first exposed to language A, and a child’s exposure to languages A and B
is fairly regular, i.e. the child is addressed in both languages almost every day” (De Houwer,
1990: 3). The history of BFLA is impressively long considering the strict\footnote{The original definition of BFLA proposed by Meisel included only children addressed in two
languages immediately after birth. The definition proposed by De Houwer (1990) is its “weaker”
version.} criteria according to which only the studies based on investigating children who are addressed in two
languages so soon after birth can be included. The major aim of the research carried out
within the frame of BFLA is determining whether the developmental path and time course of
language development in BFL learners is different from that of children learning only one
language, and if it is, then how different and why. Some other underlying issues discussed by
BFLA include determining any possible slowing down effect on the rate of general language
development, and also finding out whether “exposure to two languages simultaneously
influences the pattern of development so that it differs from that observed in monolingual
learners” (Genesee, 2005: 23). The present study does not compare the development of the
Polish gender system in bilingual and monolingual children. Nonetheless, the comparative
analyses of the participants’ mixed and monolingual utterances indicate that the individual
mechanism adopted by each child in learning grammatical gender is similar in both contexts:
the mixed and the monolingual.
BFLA studies support a view that although bilingual children are able to differentiate between the two languages (syntactic and pragmatic level) from early on, the two systems are believed to be separate, albeit not autonomous. Consequently, some cross-linguistic influence is very likely to take place, and therefore, the identification of the factors that determine cross-linguistic interference, as well as its locus and direction, are major issues in research on bilingual children. The BFLA researchers are not unanimous with regard to the source of such interference. Some claim that it should be attributed to externally controlled mechanisms, and thus, dominance as the main external factor is responsible for the cross-linguistic influence from the dominant to the weaker language (see Paradis and Genesee, 1996). Others suggest that cross-linguistic effects are due to internal sources, for example the syntax/pragmatics interface (Müller, 2000). The latter view is becoming increasingly popular, as recent results show that dominance could not explain for example the object-drop phenomenon, taking place in the dominant language.

In one of the most comprehensive studies on cross-linguistic data, in which a large amount of atypical data is investigated, Döpke (2000: 2) points out that there is general "acceptance that even the earliest grammatical structures of bilingual children indicate that their two languages develop along the language-specific lines." Contributors to this reliable and up-to-date study (Lanza, Paradis, Hulk, van der Linder and others) share the interest in the cross-language influences, which, as they believe, "are not disproving of separate development hypothesis", but can be taken as "a means of tracking the cognitive processes involved in the simultaneous acquisition of two languages" (Döpke, 2000: 2). The results of this study are also in accord with this view: the children import English nouns and make agreement choices in mixed NPs, and they also show signs of cross-linguistic mixing of morphological material in words such as *kisac* (English *to kiss* + Polish *calowac* ‘to kiss’). There is no evidence of Polish nouns imported into English utterances, which would act in the opposite way, i.e. strip Polish nouns of inflection. That shows that to some extent, the two systems develop separately, and to some extent, jointly. We shall refer to the source of cross-linguistic interference in Chapter Four.

Researchers like to make predictions as to what their language data is going to show. Unfortunately, as Döpke (2000: 6) rightly points out, by making predictions we assume we have a very clear picture of the role of the child’s cognition in the language acquisition process. As there is still a lot of uncertainty about how the minds of children work with respect to language acquisition, predictions as to where cross-language influences are likely to occur are difficult to make. Perhaps it will not even be solved in the short term, so anyone who expects the battle between nativists and empiricist to end may be disappointed.
2.3 Acquisition of grammatical gender

There is an important distinction made by linguists between the terms language acquisition and language learning. Children acquire language through a subconscious process during which they are unaware of grammatical rules. They get a feel for what is, and what is not, correct. In order to acquire language, children need a source of natural communication, and their need to communicate paves the way for language acquisition to take place. Language learning, on the other hand, is the result of direct instruction in the rules of language, and is not communicative. In language learning, students have conscious knowledge of the new language, and can talk about that knowledge. It is understood that children naturally acquire their first language. Adults (post-critical period) do not naturally acquire their second language, as a number of fundamental differences appear in their rationale towards learning. The distinction between the acquisition and the learning of a language in general, also applies with reference to grammatical gender in particular. In recent years, the processing of gender information has been investigated extensively, by both psycholinguists (interested in the acquisition of gender by mono- and bilingual children), and applied linguists (interested in the learning of gender by school children and adults). Below is a review of some of the most important findings.

According to De Houwer, in late 1980s and early 1990s, the body of research was too small and many studies were methodologically too weak to work on. Since then, some attempts have been made to solve the puzzle of how children assign nouns to genders and how they are able to learn correct agreement forms and use them as productively as they do even as early as age two. In Corbett's view (1991: 82), research on language acquisition can provide a valuable and necessary insights to the way such systems work on condition that basic linguistic description of the phenomena to be investigated is available. Such description has often been regarded as lacking from the research conducted on Indo-European languages. One reason is the notorious confusion of gender and declensional type, but also failing to recognise the regularities present in gender systems (Corbett, 1991: 82). A child learning a morphologically complex language must first recognise that despite the unlimited number of noun phrases, within those noun phrases certain regularities occur. The child must first discover a pattern, i.e., “the fact that the occurrence of certain agreeing forms (on verbs, adjectives and so on) depends on the presence of nouns of a certain gender” (Corbett, 1991: 82). This approach stands in opposition to the rote-learning model of acquiring gender as a part of each noun:

First, native speakers typically make few or no mistakes in the use of gender; if the gender of every noun were remembered individually, we would expect more errors. Second, words borrowed from other languages acquired a gender, which
shows that there is a mechanism for assigning and not just remembering gender. And third, when presented with invented words, speakers give them a gender and they do so with a high degree of consistency. Thus native speakers have the ability to 'work out' the gender of a noun... Corbett (1991: 7).

A more likely route to learning a gender system is therefore believed to begin with recognising the patterns operating in a given language, rather than memorising every single noun with its gender. Unfortunately, Corbett's prediction of "we would expect more errors" is based solely on observation of native speakers' seemingly error-free use of gender. To make predictions of that sort we would need a clear-cut language acquisition model to show us the storage and retrieval mechanisms of gender-related information in the speakers' memory. Despite the fact that research in the acquisition of grammatical gender is still lacking from many languages, what emerges from the available work is that children do not follow one particular order of acquiring the distributional rules. They often depend on the morphological complexity of the language, as well as the relative strength of the extralinguistic (semantic) and interlinguistic (formal) clues available to them. §1.0.3 of this thesis outlines the rules for gender assignment based on semantic and formal information. It shows that the semantic core is always present in all languages, but assigning all nouns (animate and inanimate) to genders requires access to the formal criteria. The studies carried out oscillate between two theoretical positions (Pérez-Pereira, 1991b: 4):

a. gender differentiation is established on the basis of semantic features coming from extralinguistic reality (natural gender theory) - children primarily attribute the gender of words on the basis of the semantics (Mulford, 1983)
b. gender is a phenomenon of the linguistic system - agreement is the essence of gender; children are able to recognize that e.g., nouns with a particular ending co-occur with other parts of speech such as pronouns or articles; morphological and syntactic data is the most important (see Karmiloff-Smith, 1979, Maratsos, 1980, Levy, 1983a)

There is no doubt that the semantic information present in animate nouns can support children's choice of the agreement forms. Children are said to be able to extract information from sexual dimorphism from the age of approximately 2;6, which is also when they acquire gender identity and can identify themselves as belonging to one of the gender groups. Moreover, according to the semantic primacy hypothesis, "the child first pays attention to the semantic properties of nouns, i.e. semantic gender rules should be acquired before formal rules" (Müller, 1990: 208). Nevertheless, the majority of studies on the

23 The difference in form between individuals of different sex in the same species
acquisition of gender support the approach according to which children pay far more attention to intralinguistic gender clues from early on. Smoczyńska (1985b: 644-8) reports on Polish children who acquired grammatical gender distinctions by the age of two. Slobin’s (1973) concept of grammaticalizable notions representing animate nouns is the basis for the so-called semantic learning, when a child combines nonreferential units with nouns on the basis on the semantic features (+/- male) of the nouns. Importantly, whenever relevant semantic features cannot be found, the child is believed to make a step towards using the formal, i.e. morphological and phonological, markers. Such procedure is referred to as form-related learning. Müller (1990: 208) also mentions what is known as distributional learning, when after failing to find phonological clues to the choice of a gender marker, a child tries to differentiate the markers on the basis of elements that systematically co-occur with the noun used and creates a paradigm of forms.

Scholars investigating aspects related to grammatical gender devoted most of their time to morphologically complex languages. Studies conducted in 1970s and 1980s discussed the interplay of intralinguistic and extralinguistic clues. In her study with Russian preschool children, Popova (1973) investigated the conditions promoting the acquisition of morphology in general, and the significance of the morphologically obligatory agreement in particular. The main area of interest in this study was determining the gender of the past tense verb agreement, but children’s orientation in creating interdependent word combinations was tested as well. The experiment was divided into three stages. In stage one, Popova tried to establish which features appear in the agreement between nouns and verbs in the past tense by choosing series of verbs. The findings obtained from the first experiment show that generalizations of one gender (feminine or masculine) that appear in a child’s speech based on a certain number of nouns may hinder orienting to the form of the word. They led to the next experiment (stage two), in which Popova examined conditions that favoured a child’s orientation to a particular form in the process of acquisition of correct agreement. In stage two, the goal was to teach children certain noun-verbs combinations, and by doing so, single out endings and generalize them in the agreement of nouns and past tense verbs. Based on stage one experiment, Popova distinguished several stages of acquisition of gender agreement: Stage I – predominance of feminine, Stage II – mostly masculine, Stage III – gender confusion, Stage IV- correct agreement. It is believed that feminine gender is preferred in the early stages of the acquisition of Russian, and it forms a connection of the first type, as opposed to a connection of the second type created by the vocabulary of masculine forms of past tense verbs. “The stage of confusion” (III) is likely to be the most intriguing stage when children begin to solve the puzzle and sort out the possible regularities governing the agreement. Only at this stage are children able to correct themselves. The
findings showed that not all children were able to learn correct gender agreement despite the teaching, and those who did it fastest were in the most creative stage, the so-called “stage of confusion” (III). Popova’s second experiment showed that “the duration of predominance of masculine or feminine gender is caused by generalisation of the agreement connections which become fixed in the child’s speech” (Popova, 1973: 276). The conditions favouring the choice of one form over the other are not accidental. Children’s generalisations are reinforced by the fact that some of them are, in fact, correct. Teaching of the proper agreement forms to the children will not result in their error free acquisition of agreement, and “mere accumulation of verbal experience in nursery conditions, without a special organization of the linguistic material which would facilitate orientation to the phonetic aspects of words, does not accelerate the acquisition of gender agreement (Popova, 1973: 277). This conclusion led Popova to designing the third experiment, which involved teaching and practical activities, during which children were expected to form correct agreement forms. Depending on their answer, they were either allowed to carry out a given activity, or not. During this stage, Popova observed that, apart from hindering the orientation to the form of the word, gender generalizations were also an obstacle in repetition of the correct word form after the experimenter. Therefore, the third experiment was designed. It was aimed at switching children from generalizations of one gender, and involved teaching past tense verb forms contrary to the one predominant in the child’s system. It was done so in order to avoid fixation of forms, and it resulted in the replacement of the old connections with the new ones. This stage was soon followed by “gender confusion” stage, which was then followed by children’s use of both feminine and masculine forms in past tense verb forms, enabling the experimented to teach them more words with contrasting genders. This eventually led to the acquisition of correct agreement.

Popova’s experiment showed that “the change in the character of agreement is directly dependent on the noun structure” (Popova, 1973: 272). In other words, the noun ending directly influences the nature of the agreement. Popova showed that Russian children rely on the morphological features (intralinguistic clues) of the noun more than on the semantic features (extralinguistic clues) when building sentences with past tense verbs. Similar results for Russian have been reported more recently by Chirsheva (2009), who shows that gender in Russian is assigned according to a complex set of semantic, phonological and morphological rules, and that “two main strategies of gender assignment are observed: semantic and phonological analogy” (Chirsheva, 2009: 85). Popova’s main findings were: two types of connections are established in child’s speech: noun + feminine verb form, and noun + masculine verb form (importantly, those connections often become fixed in children’s speech due to some of them being correct); children’s orientation towards
the phonetic features of words is supported by special organization of linguistic activity (orientation to sound is a crucial aspect of the success of the entire activity). Popova also observed that the basis for correct agreement and for moving from one stage to the other is the child’s orientation to the formal features of the noun. These findings were confirmed by Levy (1983a), Pérez-Pereira (1991a), Chirsheva (2009) and others. This study also supports the claim that children have a mechanism in which the generalisation stage is adopted as a strategy to increase the likelihood of producing a correct gender assignment, and that this generalisation is made based on the formal features of the noun.

In her papers, Levy argued against the theoretical motivation underlying a stage model for language development proposed by Gleitman (1981). According to this model, early grammars are mainly of a semantic nature. Gleitman’s maturationally determined stage process proposed as an egg/tadpole/frog hypothesis stated that the change enabling the child to develop a formal representational system allows for the development of a formal grammar later in the course of language acquisition, while the earliest representational forms are conceptual. According to this theory, children start with a semantic-based grammar, and move to the formal-based grammar with the help of the maturational process. According to Levy, however, the initial stage in the acquisition of grammatical gender is certainly not solely semantic-based. In search of the answer, she focused on referential pronouns, inflected verb forms and animacy feature in a variety of languages, and enumerated four routes which children may choose to take in learning gender distinctions: a) a semantically-based route, by which a child comes up with semantically based generalizations for animate nouns, and extends her assumptions to inanimate nouns; b) a formal route, in which she distinguishes formal (syntactic and phonological) clues; c) no particular route, which would mean no typical overgeneralizations; d) ad hoc learning, in which no rule-governed behaviour occurs. Levy’s observations of the speech of her own son as well as a brief examination of gender systems in a number of languages have shown that although “it is reasonable to expect considerable individual differences among children” (Levy, 1983: 89), a similar pattern of gender acquisition for various languages is found. She found that children not only discover formal-distributional patterns (particularly when they are systematic), but also master phonological properties of words early on, which means they possess clues to inflectional patterns also quite early. Another of Levy’s findings also shows that more errors occur whenever semantic rules are needed to determine the appropriate choice of the form (e.g. masculine personal versus non-masculine personal in Polish). These results refer to

---

26 In recent years, studies on linguistic relativity and gender have become increasingly important. A number of experiments by Boroditsky et al (2007) show that linguistic information helps shape semantic representations throughout development.
languages with complex gender systems, rather than languages such as English, where
gender system is of a purely semantic nature.

Henzl’s (1975) experiment with Czech speaking toddlers is noteworthy for her
findings concerning the order of gender assimilation and the children’s attempts to derive
grammatical rules for assigning gender from the input. For her experiment, Henzl recorded
three Czech-speaking children, aged 1;9, 2;10 and 3;8. It involved a series of tests, which
were carried out every three months, and was typically an elicitation experiment, through
which Henzl hoped to examine both simple gender agreement in attributive adjectives and
verbs, and two-way agreement in possessive adjectives (similarly to Polish, gender in Czech
is based on the phonological clues on the noun, while the distribution of genders is:
masculine 43%, feminine 38%, neuter 19%). The elicitation material included 13 real Czech
nouns and 16 nonce words. Each word was represented graphically (nonce words had
abstract representations) in three different ways: in a large size, in a small size and in a
falling position. The experimenter showed pairs of cards to each child, introducing the
object, and asking about it. Attributive agreement was elicited by questions such as: “What is
the name of the cat here”, whereas verbal agreement was elicited by asking: “What’s
happened?” while pointing for example to a drawing in which a cat fell down. Possessive
adjectives were elicited by asking “Whose X is that?”.

The score indicates that a clear-cut phonological form enables Czech children to
classify gender correctly. In addition, similarly to Russian children reported by Popova
(1973), Czech-speaking children tend to make generalisations reinforcing the predominance
of one agreement form over the other. The reasons seem similar: some connections become
fixed in child’s speech, and children’s choice is reinforced in the process of communication.
Henzl underlines the period of regularizing the first gender assignments, and places it
between the time of undifferentiated usage and the productive usage of Czech inflectional
morphology available to young learners (1975: 192). Henzl’s study also presents some
established principles in the marking of gender. Firstly, phonological clues prove more
productive whenever they fit the representative gender categories. Secondly, many
structurally ambiguous words, (monosyllabic in particular), trigger neuter gender. Thirdly,
children avoid an unsatisfactory decision by regularising the input word, e.g., by giving it a
clear-cut ending of their own choice. Similar mechanisms have been reported by
Smoczyńska (1985b) for Polish children, who tend to regularize irregular feminine nouns by:
adding a feminine ending -a and declining the noun according to the regular feminine
declensional pattern (e.g., koleja instead of kolej ‘railway(f)’); avoiding the main form and
using a diminutive form -ka instead (e.g., myszka instead of mysz ‘mouse(f)’); finally, by
treating them as masculine altogether and declining accordingly.
Pérez-Pereira (1991a) replicated an experiment introduced by Karmiloff-Smith (1979) in which the importance of the intra- and extralinguistic cues was tested. The key to this experiment (designed to overcome drawbacks found in other experiments carried out with a similar goal) was the fact that children were presented with nouns including all possible combinations of the three clues to noun gender available in Spanish (semantic, morphological and syntactic). Thus, in some items, only one of the clues was present, while in other, there were either two agreeing clues available, or two different clues were in conflict. For instance, in *un nepo* a syntactic clue is presented via the masculine determiner *un*, while the morphological clue is available via the suffix -*o*. Such a procedure allowed the investigator to observe closely the interplay of clues made available to the subjects, and consequently, to draw reliable conclusions about the relative importance of the intra- and extralinguistic clues that Spanish children may use in the process of learning Spanish gender.

The results to this study show that Spanish children master gender agreement by the age of two. Such an early and accurate acquisition of gender distinctions is attributed to the fact that Spanish has two genders and the marking of gender is more systematic and clearly differentiated. The later mastery of gender system in languages like Polish or Russian, on the other hand, is blamed on the ambiguity and unpredictability of morphological marking, e.g. the overlapping of gender and case markings. It may, therefore, seem confusing to learn from various studies that “the more extensive and productive the system of gender marking is, the easier is its learning, since it furnishes more frequent and concordant information to be used by children” (Slobin, 1982, see also Mills, 1986a, Pérez-Pereira, 1991a: 585). The solution to this puzzle is Slobin’s claim that “the acquisition of linguistic systems in which there are clear and binary distinctions offering formal criteria for suffix addition is easier than the acquisition of those lacking such characteristics” (1982: 56, after Pérez-Pereira, 1991a: 587).

In comparison with English, German has a more productive and extensive system of gender marking which involves various parts of speech. In comparison with Spanish, however, the German gender system is more ambiguous and less systematic. Spanish has two genders, while Polish has three. However, the important difference lies in the clarity, regularity and productivity of a gender system. Another important finding in the Spanish study shows that items with masculine clues are more readily recognised by Spanish children than items with feminine clues. Similar suggestions are made by Lehečková (2000: 746) regarding Czech, where correct masculine nouns are learned earlier, although in the first stage of the acquisition of agreement, Czech children use the feminine gender agreement forms more productively. It is possible that although masculine is easier to acquire, in various languages the sequence of the agreement forms learned will vary depending on the salience and frequency of use. For instance, longitudinal work on Hebrew (Levy, 1983a) supports Pérez-
Pérez-Pereira’s experiment shows that Spanish children mainly rely on intralinguistic rather than extralinguistic information when establishing gender agreement. Older children pay more attention to syntactic information than to morpho-phonological information in items where there is a conflict between the clues. For them, items providing syntactic clues are clearer than items with extralinguistic clues. In other words, the presence of the intralinguistic information plays a significant role in the learning process.

Although Van der Velde’s (2004) study\textsuperscript{27} investigates monolingual children between 3-6, it is worth mentioning, as it discusses the concept of default gender, i.e. the mechanism of overgeneralizing one gender over another. In an experimental elicitation task, Dutch and French children are expected to produce article+noun sequences in both isolated contexts and inside clauses. Van der Velde’s results show that children’s acquisition of gender differs in these two languages. Dutch children initially use the determiner *de* as the default gender with both neuter and non-neuter singular nouns, while French children make hardly any gender errors at all, i.e., do not use any gender in the default way. Dutch children are reported to be using the correct neuter definite determiner *het* very slowly and only occasionally. As to French children, while hardly any gender errors are found in their speech, it has been suggested that the correct use of determiners in the early stages does not necessarily imply that the children have acquired the intricacies of the gender feature. As Hulk (2007: 181) puts it: “It may be the case that initially the selection of the gender is solely determined by phonological shape of the noun and/or by probabilistic correlation”. The default character of masculine in Polish is confirmed in this thesis, but it is not a simple matter to say that it is always due to the phonology. Each of the three children studied here overuses masculine in some way, but the varying time span of such an overuse suggests that even if initially their choice is phonology-based, the change from overusing masculine to feminine may be, in fact, morphology-based, i.e., the reliance on the form of the noun may be transferred to the reliance on the interrelation between the noun and the modifier.

For reasons of space, we abstract away here from the important literature on the acquisition of gender by adults/L2 speakers and will continue with a brief summary of more recent work on the acquisition of gender by children (Table 9 below illustrates the amount of work done in the acquisition of gender since late 1940s). A number of researchers have embraced recent models of language acquisition, and with the use of e.g. neurolinguistic testing, have suggested ways in which gender systems may be operating. Friederici et al. (1999) investigate the processing of grammatical gender during language comprehension.

\textsuperscript{27} After Hulk (2007).
mainly focusing on gender priming. The authors review behavioural data in priming experiments and show that regardless of language and task, strong inhibition effects exist. Behavioural experiments also show interaction of gender priming and semantic priming effects. The authors explain the interaction on the basis of the ERP data “as taking place at a late stage of processing” (Friederici, 1999: 482). Their results posit that “during normal language comprehension gender information does not preselect particular (gender agreeing) lexical candidates, but that a postlexical checking mechanism evaluates the gender congruency of the incoming element” (Friederici, 1999: 482). We should note that this suggests the presence of lexicon containing links between nouns and genders. These findings also suggest “a general independence of syntactic and semantic aspects during the early stage of comprehension” (Friederici, 1999: 482). We could interpret such a conclusion as a continuation of the observations made in earlier studies (see e.g., Henzl 1975), which suggested that children initially rely on the least ambiguous forms, rather than the most frequent ones. In other words, formal clues have rightly been reported to guide the children more reliably than the semantic clues (see Karmiloff-Smith, 1979, Maratsos, 1980, Levy, 1983a). The gender priming study also indicates that the interaction between the syntactic and semantic aspects takes place at a later stage, which confirms the observations of those researchers who reported data where children made fewer gender mapping errors, since they understood the relationship between the natural and grammatical gender.

The results of studies by Gordon (1985), Gathercole (1985), Katz, Bakos, Macnamamara (1974) show that young children have the ability to distinguish different kinds of words taking account of their distributional patterns. Kempe (2003) showed that noun diminutivization can support the correct gender assignment by young Russian speakers. Findings in experimental studies of Brazilian child language acquisition reported by Name & Correa (2002 & 2003), interpreted from the theoretical framework of Chomsky’s (1995 & 1999) minimalist programme, shed light on the role of agreement syntax in learning gender and number grammar at the early stage of language development. Johnson (2005) investigates Dutch toddlers’ perception of definite determiners, and the agreement between determiners and nouns. These are but a handful of studies available to linguists looking into the process of the gender acquisition in child speech. We now turn to studies available to researchers working on Polish.

Compared to other languages, studies dealing with the acquisition of Polish grammatical gender by monolingual children are nearly non-existent, and the situation involving bilingual children who learn Polish as one of their languages is even worse – there are no studies of this kind whatsoever. Most of the research conducted (dating back between 28 Gender priming involves introducing a lexical element before testing it in an experiment.
70-40 years ago) contains detailed analyses of child Polish ranging from behaviouristic observations to phonological analyses. However, the children's acquisition of Polish gender system is notoriously confused with the acquisition of the inflectional morphology, which makes any interpretation impossible. For example, Smoczyńska’s (1985b) most detailed account of the acquisition of Polish so far provides some information on the order of the acquisition of inflectional endings, but the author seems to be mistakenly treating the acquisition of the declensional classes as the recognition of the patterns which the children need to learn to develop their gender system. Smoczyńska claims that the ability to use specific case endings almost entirely correctly from the early age can be assigned to the early mastery of grammatical gender in the singular. Unfortunately, declensional patterns comprise only one side of the coin. Let us consider the noun *tata* ‘daddy’, whose pattern of declension is determined by its -a ending (see §1.1.2.2). The child who declines *tata* correctly according to the feminine paradigm has learned to decline a noun with the ending -a according to the appropriate declension class. It does not mean that the child has discovered the natural gender rule, which *de facto* allocates *tata* to the masculine gender. The fact that the children assign declensional endings properly does not prove that they have learned the gender. Nouns in Polish must be placed in agreement with other parts of speech, and the main deciding feature here is grammatical gender. This fact should not be confused with the acquisition of case endings. Any relationships between the case system and the gender system should be assumed only with evidence of agreement confirming children’s ability to see the regularity in placing nouns and modifiers in agreement in accordance with the rules of the gender system. At the moment, the main focus of investigation in child Polish seems to be the acquisition of the lexicon (Haman, 2002) and the development of the inflectional paradigms in Polish monolingual speakers (Dąbrowska, 2001, 2006, 2008). A small number of researchers investigating Polish are also currently working within the usage-based approach (see §2.1) with such scholars as Elena Lieven and Michael Tomasello. Krajewski’s (2005) work on the role of grammatical gender in the acquisition of noun inflection is the only study which, apart from discussing the Polish case system, also discusses gender system as a separate topic. According to Krajewski, “the acquisition of the declensional system depends to a large extent on individual differences and possibly on varying input rather than on innate constraints” (Krajewski, 2005). Such statements should be made carefully, as they imply no inherent order of acquiring the case system, which was suggested 20 years earlier by Smoczyńska (1985b) (as discussed below). The source of data

---

29 Max Planck Institute for Evolutionary Anthropology, Department of Developmental and Comparative Psychology, Leipzig
30 Krajewski’s work is based on data obtained from a child involved in data collection in Kraków in 1960s.
employed by Smoczyńska is considerably larger than Krajewski’s, whose study (as he
admits) should be treated as a preliminary analysis and should be replicated on a larger
corpus to enable longitudinal analysis. Despite the time that has elapsed from Smoczyńska
study, we are more likely to rely on her conclusions here. Krajewski also claims that some
gender classes are available from the onset of the acquisition of noun inflection. To be clear
about the terminology here, as has been explained in §1.0.3 of this thesis, the term gender
classes can be used to mean noun classes or genders. However, regardless of the
terminology, the availability of noun/gender classes does not prove the ability to recognise
the working of the gender system.

Table 9. Research in the acquisition of gender

<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Language</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>Gvozdev</td>
<td>Russian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1957</td>
<td>Bogojavlenskij</td>
<td>Russian</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1959</td>
<td>Rute-Dravina</td>
<td>Latvian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1966</td>
<td>Slobin</td>
<td>English</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1968, 1972</td>
<td>Pacesova</td>
<td>Czech</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1970</td>
<td>Omar</td>
<td>Arabic</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1973</td>
<td>Pupova</td>
<td>Russian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1973</td>
<td>Zakharowa</td>
<td>Russian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1975</td>
<td>Sadek-Kiraithe-Villarreal</td>
<td>Spanish+English</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1975</td>
<td>Henzl</td>
<td>Czech</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1975</td>
<td>Radulovic</td>
<td>Slovene</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1978</td>
<td>MacWhinney</td>
<td>German</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1979</td>
<td>Karmiloff-Smith</td>
<td>French</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1983, 1988</td>
<td>Levy</td>
<td>Hebrew</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1983</td>
<td>Mullford</td>
<td>Icelandic</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1983</td>
<td>Taeschner</td>
<td>German+Italian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1984</td>
<td>Hernandez Pina</td>
<td>Spanish</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1985</td>
<td>Clark</td>
<td>French</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1985</td>
<td>Smoczyńska</td>
<td>Polish</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1985</td>
<td>Berman</td>
<td>Hebrew</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1986</td>
<td>Mills</td>
<td>English</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1986</td>
<td>Mills</td>
<td>English, German</td>
<td>both</td>
</tr>
<tr>
<td>1987</td>
<td>De Houwer</td>
<td>Dutch</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1989</td>
<td>Taraban-McDonald-MacWhinney</td>
<td>German</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1990</td>
<td>De Houwer</td>
<td>English+Dutch</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1990</td>
<td>Phunkett-Strömqvist</td>
<td>Scandinavian lgs</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1991</td>
<td>Pérez-Pereir</td>
<td>Spanish</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>1992</td>
<td>Connors</td>
<td>French</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1992</td>
<td>Pizzuto</td>
<td>Italian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1992, 1995</td>
<td>Andersson</td>
<td>Swedish</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1994</td>
<td>Koehn</td>
<td>French+German</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1994</td>
<td>Müller</td>
<td>French+German</td>
<td>longitudinal</td>
</tr>
<tr>
<td>1997</td>
<td>Dressler</td>
<td>Polish</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2000</td>
<td>Wegen</td>
<td>German</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2000</td>
<td>Müller</td>
<td>French+German</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2000</td>
<td>Lobecková</td>
<td>Czech</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2001</td>
<td>Debrowska</td>
<td>Polish</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2001</td>
<td>Guillemin-Grosjean</td>
<td>English-French</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>2001</td>
<td>Gathercole</td>
<td>Welsh</td>
<td>cross-sectional</td>
</tr>
<tr>
<td>2003</td>
<td>Kempe</td>
<td>Russian</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2005</td>
<td>Johnson</td>
<td>Dutch</td>
<td>longitudinal</td>
</tr>
<tr>
<td>2004</td>
<td>Bohnacker</td>
<td>Swedish</td>
<td>longitudinal</td>
</tr>
</tbody>
</table>
According to the data obtained from the Kraków database, Polish-speaking children learn gender distinctions in the singular (m, f, n) at a very early age. Such precocious acquisition allows for the simultaneous emergence of some early inflectional endings in accordance with the gender of the noun, and, as a result, any erroneous agreement forms are limited to a very short period. Moreover, those errors “should be attributed to the lack of knowledge of possible differentiation of the forms rather than of that concerning the noun gender” (Smoczyńska, 1985a: 645). The opposite phenomenon is the delayed acquisition of masculine personal/non-masculine personal distinctions in the plural. Since gender division in the plural does not mirror the one in the singular, children take longer to learn the extra number of separate rules governing the plural. Masculine personal nouns have a specific and highly irregular distribution of endings in the nominative plural. Not surprisingly, errors in this area are observed in the later period of the acquisition of Polish, namely around the age of 3-4. Children’s delayed acquisition of the masculine personal/non-masculine personal distinction is one of the most intriguing areas of the acquisition of gender, with the accusative plural at the centre of attention. The accusative plural does not match the rules of the singular. In the plural, masculine personal nouns follow accusative-genitive, while non-masculine personal nouns follow accusative-nominative. In the singular, the case syncretism of accusative-genitive takes place for masculine animate only. Consequently, any masculine animate nonpersonal noun used in the plural is likely to be a source of confusion. Children may remember that, e.g., the names of animals in the singular have the same ending for the accusative and the genitive. When they start using the plural, they are likely to continue applying this rule without realising that animals do not belong to the masculine personal group of nouns, and that the accusative should match the nominative. Solving a puzzle of that kind is likely to take time and effort. Therefore, Polish-speaking children “tend to leave it” until most of other rules are at least familiar to them. According to Smoczyńska (1985a: 467), children’s late acquisition of the masculine personal-non-masculine personal distinction in the plural is due to the formal complexity and the rarity of male groups in the child’s experience. Levy (1983a: 83) casts doubt on this explanation claiming that even preschoolers are familiar with such vocabulary items as firemen, soldiers, policemen, brothers, and male friends and that the morphological complexity cannot be a sufficient explanation, because Polish children master equally complex parts of the system much earlier. Levy believes that masculine personal/non-masculine personal distinction in the Polish plural “is a case in which an understanding of the semantic notion is crucial for the child to be able to work out the formal intricacies of the system of agreement for plural nouns” (1983a: 83). The data gathered for this study shows that the three children had a number of occasions to familiarize themselves with male representatives such as strażacy
firemen(M.PERS), chłopcy boys(M.PERS), panowie men(M.PERS) etc. between the age 2;4-4;1, via book reading, story telling and from other sources. Nevertheless, even a cursory glance at the data shows minimal traces of masculine personal agreement of any form, correct or incorrect. None of the children attempted to use masculine personal productively during the course of recording. Moreover, much later, when the children turned 4;5-5, observations made during friendly post-recording visits showed no interest whatsoever from the children to disambiguate masculine personal. Levy’s claim regarding the need to understand the semantic notion to untangle the formal intricacies of the plural agreement is sound, but 5-year-old preschoolers are surely able to understand the semantic notion of masculine personal nouns in question. It seems that there is much more to their learning of masculine personal/non-masculine personal than semantics, and the likely candidate is the mechanism suggested by Popova (1973) - the generalisations of the agreement connections become fixed, and are reinforced by the fact that some of them are correct.

Before we proceed with the description of the methodology (Chapter Three), and the analysis of the data (Chapter Four), let us present the main hypotheses of this thesis. Firstly, this thesis is expected to uncover details about learning strategies employed by the Polish-English children when learning grammatical and natural gender. Gender seems to be one of the features of language that children do not imitate. Moreover, it cannot be innate, as children would produce fewer errors, and the learning strategies would be more transparent. The situation is different: children produce gender errors, and their routes to learning gender are not easy to identify. On the contrary, the strategies can be highly individual. It is likely that rather than being equipped with some innate knowledge of grammatical gender they apply certain learning strategies. Secondly, based on the reported fact that bilingual children separate their two languages very early on it can be hypothesised that the speech data of Polish-English children will show little evidence of cross-influence between the Polish and the English gender system. Thirdly, based on other reports investigating the acquisition of grammatical gender, we expect the children to use formal criteria before semantic criteria. This expectation raises an additional question: if formal criteria are used before semantic criteria, is the animacy feature going to be compromised? If so, for how long? Finally, following Piaget’s concept of consistency in errors it is believed here that gender assignment errors are a source of significant information about the development of gender system in each child, and they can uncover patterns which point to gender learning strategies. This study is also hoped to provide evidence of how bilingual children organise their linguistic knowledge when faced with two very different gender systems. Such data is severely understudied in bilingual child language area. The insights provided by the Polish-English data will contribute to building a more complete picture.
2.4 Conclusions

Research in developmental psychology, linguistics and cognitive science of the past 20 years suggests that children acquire languages without the aid of any hypothesised universal grammar. Two fundamental points are made: (1) children have at their disposal much more powerful learning mechanisms than simple association and blind induction; (2) there exist plausible and rigorous theories of language that characterize adult linguistic competence in more child oriented terms than does generative grammar, which makes the endpoint of language acquisition seem much closer (Tomasello, 2003: 3). Research on child language development confirms that children are not pre-wired with some innate knowledge of grammar. Tomasello discusses Single and Dual Process Models to show that a set of cognitive, social-cognitive and learning skills help children in understanding communicative intentions, as well as in discovering patterns among utterances and morphological markers. A dual process (mentioned in §2.1) predicts that children "learn irregular morphology in a manner similar to that proposed by the connectionists\textsuperscript{31}, but regular morphology is different because it participates in some way in the innate and abstract rule system of universal grammar" (Tomasello, 2003: 237). The rule-based process in which irregular forms are learned is expected to be insensitive to factors affecting the learning and the use of the items not participating in the rule, i.e. type/token frequency and semantic/phonological similarity among exemplars. In all, in dual process frequency, regular items are insensitive to semantic/phonological similarity and frequency, whereas irregular items are sensitive to all of these factors. In the single process model, on the other hand, abstract schemas built by children during the learning process are tied to the learning processes. The schemas underlying productivity simply reject symbolic rules immune to normal learning processes. This model posits that

words enter the child’s lexicon with a certain strength based on token frequency. High token frequency (strength) of an item (for example, \textit{was} as a past tense) enables it to resist assimilation to any generalized schema. Words similar in semantic and/or phonological form cluster into schemas. The productivity of a schema is a function of (1) the singularity among exemplars (such as in terms of semantic or phonological properties); and (2) its type frequency in terms of the number of different lexemes with which it has been used. (Tomasello, 2003: 238)

To illustrate the functioning of the single process model, Tomasello discusses German plural -\textit{s}, English past tense -\textit{ed}, and Polish genitive. German -\textit{s} both has few phonological restrictions, but it is not entirely open, because of its use with a limited set of nouns. It is thus a "default" (it has few restrictions, but is also a learned schema). The English past tense

\textsuperscript{31}a totally data-driven single-mechanism theory that does not depend on any \textit{a priori} formal linguistic theory
marker is quite open phonologically and is used very frequently, which makes it more than a
default (it is referred to as “regular”). Polish genitive has three endings, whose distribution is
mostly determined by gender (also phonological, morphological and semantic factors). It has
no default ending in masculine. Dąbrowska (2001) examined Polish child language data and
found out that overgeneralisation errors involve irregular forms, rather than defaults.
Tomasello believes that this finding shows that children are not presented with a pattern of
input suggesting a default form. Their early acquisition of Polish genitive supports the view
that children are not inclined to impose a default form.

Studies which investigate gender acquisition in second language learners, are more
frequent than studies describing mono- or bilingual child acquisition of gender. Among the
studies on the acquisition of gender there are very few that investigate this issue in the
speech of bilingual first language learners. Only a few have studied children as young as the
three participants of this study, and many focus on school children, which does not allow
reliable age-to-age comparison between them and this study. Those that analyse the
acquisition of gender by adults are not comparable at all, as the acquisition context is entirely
different for those two groups. Thorough research of published and unpublished materials
reveals that among studies that discuss bilingual acquisition of gender there are none
including Polish as one of the language. Smoczyńska’s (1985a) description of the
monolingual acquisition of Polish mentions the development of the gender system, but only
as a part of a general description of the morphosyntactic development in Polish children.
Older monographs (Brenstein-Pfanhauser, 1930; Zarębina, 1965) that discuss language
development of Polish children are outdated and too general to provide detailed analyses of
gender and gender agreement. In this quite impoverished environment, there certainly is a
large gap to fill in as far as data collection and data analysis involving Polish child language
are concerned. This study attempts to fill the gap. The next chapter discusses details
underlying the methodology applied to test them, as well as the general metadata pertaining
to this project.
Chapter Three: Methodology

3.0 Introduction

This chapter presents the reader with detailed information on the informants and the recruitment criteria, and describes the research methods employed in this study. The §3.1 explains the recruitment criteria, such as the age of the informants, or the communication patterns at home. It also presents the history of the children’s input and the places of residence. The second section discusses language growth measuring tools, and provides a brief discussion of the status of the morpheme, which is linked directly to the choice of the measuring tool used here. This section also describes the data collection technique, the transcribing rules, and the system of coding the data for morphological information.

3.1 Recruitment

This section discusses the reasons that guided the investigator in the process of selecting the appropriate age group for this study. It used to be suggested that when it comes to speech development bilingual children are slower than monolingual children. This view is no longer accepted. Despite varying data, the general understanding is that bilingual children go through the same stages of linguistic development as monolingual children: babbling, one-word, two-word, multiword, and multiclause. Therefore, “there is no reason to believe that the underlying principles and mechanisms of language development (in bilinguals) are qualitatively different from those used by monolinguals” (Meisel, 1986: 64), and a bilingual child whose speech is somewhat delayed should be treated exactly like a monolingual child. A good comparison has been provided by Brown (1973: 270) who conducted research on three English-speaking children (Eva, Sarah and Adam). This study shows how differently in terms of timing monolingual children learn their mother tongue. Figure 1 illustrates this timing difference.
The present progressive constructions appear in Eve’s speech around the age of 1;10, in Adam’s at the age of 2;6, in Sarah’s around the age of 2;11. Eve could be called an early-talker in comparison to Adam and Sarah, but does that make them late-talkers or is Eve’s speech unusually early? Were she bilingual, would Eve begin to speak later? According to Brown (1973), while the order of development of what he calls “the fourteen English morphemes” is quite constant, the rate of development varies widely, which can be observed when comparing the ages of the three children with their MLU (Mean Length of Utterance) stages. Importantly, bilingual children vary just as monolingual children do, regardless of which language they are learning. Due to the lack of any earlier research involving Polish-English bilinguals, the author cannot point to any comparative data presenting language development of monolingual Polish and bilingual Polish-English children. Still, research in the speech of monolingual Polish children indicates a certain order and time of linguistic development. This reported order served the investigator as a preliminary guideline for selecting the age of the participants for her study.

Although no inflectional forms are used productively at the one-word stage (1;6), Polish children start using case endings before the age of 2, and the initial contrast often is between nominative and genitive (e.g. *miś* ‘teddy-NOM’ versus *misia* ‘teddy-GEN’, Smoczyńska, 1985b). Children differ from each other in terms of when the two-word stage

---

32 MLU is computed by dividing the total number of morphemes in an utterance by the total number of utterances. The procedure will be discussed in detail §3.3 of this chapter.
starts relative to their age. If on average they start using one-word structures actively around the age of 1;6, an assumption can be made that the two-word stage is likely to begin between the age of 2;0-2;2. The first level of the two-word stage consists of *frozen phrases*, i.e. unmodified phrases that the child has memorised (e.g., *daj to ‘give-IMP this’*), or phrases in which the child learns to replace one element only (most often a noun, e.g. *mama da ‘mummy give’*). Morphological development begins around 2;3-2;5 (in the third month of the two-word stage). As the main indication of the morphological development, Smoczyńska (1985b) lists cases in order of their emergence in the children’s speech. As has been explained in §2.3 (Chapter Two), the study has little value with respect to the acquisition of grammatical gender, since the author treats the acquisition of inflection as a reflection of the acquisition of grammatical gender. Nonetheless, the analysis indicates that in morphologically complex languages such as Polish the awareness of the morphosyntactic features starts in the second year of life, and their development lasts on average until the age of 4;0-5;0. The investigator had every trust that within the time granted for recordings the children would produce enough data for her to be able to identify all the important gender distinctions, such as male/female, animate/inanimate, human/nonhuman, or masculine personal/non-masculine personal.

Let us now discuss the recruitment process in more detail. In her well-known book on the acquisition of English and Dutch, De Houwer (1990:30) highlights the significance concerning the choice of subjects: “In a cross-sectional study, the subjects must be fully matched for everything except age and/or level of language development. If not, one is again comparing apples and pears”. Any fieldwork-based research, particularly a cross-sectional study, relies on accurate recruitment, since the reliability and the accuracy of the investigation depend on it. The requirements of this project dictated that the subjects were at a similar linguistic stage for the emerging gender systems to be comparable; hence, a metric for that comparison had to be established. A careful review of the research in child language acquisition has shown that the chronological age and the MLU are two widely and conventionally applied metric types. The same method has been applied here. Although the children selected for the present study were matched with regard to their chronological age, the initial MLU differed hugely between the two boys and the girl. At the age of 29 months, the MLU of both boys equalled 1;8, whereas the girl’s MLU was 1;1 (she managed to catch up with the boys around the age of 33 months). The children’s MLU will be discussed further on. The subjects’ expected age range between 22-28 months was justified by the

---

33 The initial distinction between nominative and genitive is typically followed by the emergence of VOC sg., and NOM/ACC pl. Next are: INS sg, LOC sg, DAT sg (mixed with GEN sg), INS pl, DAT pl, GEN pl, VOC pl.
reports for both Polish and English children. As explained above, active morphological
development for Polish children starts around the age of two. As to English children, the
literature provides scholars with average guidelines as to which language stage children
achieve at what age. In the three children Braine looked at, after the first word combinations
(1;7-1;8), children showed a sudden upsurge around 2;0-2;2, which was associated with "a
marked increase in the structural complexity of utterances" (Braine, 1963:2). Peters' view on
the early two-word stage is accurate and supports the age range chosen for this project:

By 20-24 months, when children are becoming aware of grammatical functors, they
are well along in developing two major kinds of knowledge, which will help them in
this task. First, their familiarity with the prosodic structure of the language they have
been hearing [...]. Second, their expanding awareness of the sorts of functions
language can accomplish leads them to look for the linguistic means to express these
functions; [...]. Exactly how a given learner proceeds seems to vary, depending on an
interaction between the kinds of linguistic information the child is predisposed to pay
attention to, and the prosodic and morphosyntactic characteristics of the language

The aim of the recruitment stage was to select bilingual children who either just started or
were about to start using morphological markers. However, the recruitment process proved
more challenging than had been envisaged. The investigator knew none of the potential
subjects before the project started, and consequently, the children’s linguistic competence
was unknown to her as well. Furthermore, selecting children at exactly the same linguistic
stage did not guarantee that they would use gender clues creatively within the limits of the
time provided for the recording period. Thus, the extra challenge lay in the fact that the
sensitivity to morphonological clues was likely to vary across all the speakers. Consequently,
apart from the children’s matching age, three more recruitment criteria were applied for this
project: exposure patterns at home, conversation patterns between the parents, and the family
residence history. The ultimate goal of this process was to exclude children, whose
developing Polish\textsuperscript{34} would either decrease or stall in the course of data sampling. In sum, for
a dense database consisting of regular audio and video recordings, the main requirement was
a minimum of three and a maximum of five bilingual Polish-English children raised in a
mixed family, where one parent is a native speaker of Polish, and the other a native speaker
of English. One of the main criteria was that the Polish-speaking parents were brought up in
Poland and belonged to the latest wave of immigration into the UK. The requirement for the
English-speaking parents was that English was their first language regardless of its variety.
The recruitment age for children was 22-28 months, which included bilingual early-talkers as
well as late-talkers. All the families recruited for this project implemented the One-Parent-

\textsuperscript{34} Since the families were all UK based, there was a serious danger of the majority language to
overtake once the children were sent to English speaking preschools.
One-Language (OPOL) technique, which expects each parent to use their native language when addressing the child. To accommodate for the parents who speak only one of the languages, the method allows some exceptions, e.g., using a shared language at meal times. The focal recruitment area involved Greater London, Surrey, West Sussex and Hampshire.

The recruitment procedures took various forms. One of the methods involved posting the project description on the Internet discussion rooms inviting Polish mothers living in the UK to join. The discussion groups were: Matki Polki w UK ‘Polish Mothers in the UK’ Polki w Anglii ‘Polish women in England’, Polonia w Londynie i Anglii ‘Polonia in London and England’, Dwujęzyczność w rodzinie ‘Bilingualism in the family’, Forum Naukowe ‘Science forum’, Z dala od Polski ‘Away from Poland’, Kobieca Polonia ‘Women’s Polonia’, and Polscy Naukowcy za granicą ‘Polish researchers abroad’. Informants were also recruited via a letter and a poster sent to Head-teachers of a number of Polish Saturday Schools, encouraging them to help find candidates for the project. The third way of reaching potential informants was an announcement made in a number of Polish parishes around which Polish communities in the UK have long been centred. Posters were put up in churches and a few nursery schools. Finally, information about the recruitment was spread by word of mouth.

3.2 The informants

The present investigation is a context-bound study, since the quality and quantity of data available for the analysis depend on the willingness and openness of the informants, both parents and children. Initially, five mixed-nationality families were recruited. Each of them expressed considerable interest in the project and was willing to comply with the requirements. In all cases, the mothers were Polish, whereas the fathers were English-speaking. The mothers were more involved in the participation than the fathers, which is a natural consequence of the fact that the mothers were at home with the children during the recording sessions, while the fathers were at work. All the children were Polish-English by

35 the discussion groups were (all accessed regularly between Jan - June 2006):
Matki Polki w UK http://forum.gazeta.pl/forum/71.1.html?f=37418;
Polki w Anglii http://forum.gazeta.pl/forum/71.1.html?f=29377;
Polonia w Londynie i Anglii http://forum.gazeta.pl/forum/71.1.html?f=17979;
Dwujęzyczność w rodzinie http://forum.gazeta.pl/forum/71.1.html?f=37229;
Kobieca Polonia http://forum.gazeta.pl/forum/72.2.html?f=16726&w=40037626&c=40057894;
36 A full list of the Saturday Schools in London contacted by the author available in the appendix.
37 St. Joseph's Catholic Church, Guildford; St. Pius Church, Guildford; St Mary's Church, Milford; Mary of Częstochowa and St. Kazimierz Church, Devonia, London.
birth, healthy, and had not experienced prenatal or postnatal medical complications. Families were intact and of middle to upper-middle SES (Hollingshead, 1975, Four-Factor Index of Social Status). At the time of their child’s birth, mothers averaged 31.7. All mothers were university educated. Only three of the five recruited families will be described in detail here, as two of the youngest children have produced an insufficient amount of bilingual data.

3.2.1 Family i: Patrick

This family was the first to be recruited for the project. It is a four-member family living in the south of England, where Patrick was born on March 24, 2004. At the time of his birth, the mother was 33, the father was 57, and the older sibling was 1;8. Patrick’s mother, Maja, was born in Poland where she spent her childhood. She moved to the UK at the age of 12 and lived there for five years. She became more familiar with the British culture, polished her language skills, and at the age of 17 she moved back to Poland. At the age of 29 she returned to the UK to live there. Patrick’s father, John, was born in New Zealand, and moved to the UK in 1995, i.e. 9 years before Patrick was born. He is an entrepreneur, whereas the mother is a qualified teacher, but she has not been employed outside home since the birth of the first child, Ellie. Maja sometimes helps with the family publishing business. The father speaks New Zealand English, while the mother addresses the children in standard Polish. Patrick’s exposure to the two languages started from the day he was born, and it has been quite regular thenceforth. Apart from a few trips to Poland without the father, or when the father went to hospital, exposure to both languages took place virtually every day. Although the language of local environment has always been English for Patrick, his first words were in Polish, which remained the dominant language even after he had joined a kindergarten at the age of two. Between 2;5-2;10 Patrick was again at home with his Polish-speaking mother, and his father became was the main source of English for him. Patrick began attending English nursery at the age of 2;10: initially, 3-hour-sessions two days a week, extended after 3;0 to 3-hour-sessions three days a week and a 6-hour-session once a week. For the remaining weekday, he would stay at home with his mother. In total, since the age of 3;0, Patrick attended the nursery for 15 hours a week. Before the kindergarten, weekdays provided the child with more Polish than English (an average of 10 hours of Polish versus 3-4 hours of English). The exposure patterns changed after Patrick joined the kindergarten and then the nursery: on average 5 hours of English and 7 hours of Polish. On Sundays, English has always been the dominant language in this family.

Trips to Poland and visits by Polish relatives have always been an important “injection” of Polish into Patrick’s linguistic resources. Both Polish grandparents visited the
family when Patrick was born, thus the exposure to Polish in the first few weeks of his life was slightly greater than to English. In the first year of life, Patrick visited Poland for a total of 8 weeks. In his second year of life, he travelled to Poland on three occasions, spending there 11 weeks in total. In the third year of Patrick’s life, the family did not travel to Poland at all. Instead, they paid a four-week summer visit to the father’s family in New Zealand, which unquestionably enhanced even more Patrick’s exposure to English. Exactly two weeks after the family had returned, Patrick’s recording sessions for this project began, hence increasing his exposure to Polish. The Polish grandparents visited the family for two weeks at Christmas that year as well, which intensified his Polish yet again. A few short visits were paid by the family’s Polish relatives, but their influence on Patrick’s speech, albeit meaningful, is hard to measure. Between the ages 4;0 and 4;3 Patrick visited Poland once for three weeks. Table 10 below lists Patrick’s place of residence between the ages 0-4;3.

<table>
<thead>
<tr>
<th>AGE</th>
<th>COUNTRY</th>
<th>APPROX. DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-0;4</td>
<td>England</td>
<td>5 months</td>
</tr>
<tr>
<td>0;4-0;5</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>0;5-0;8</td>
<td>England</td>
<td>2.5 months</td>
</tr>
<tr>
<td>0;8-0;9</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>0;9-1;2</td>
<td>England</td>
<td>4;5 months</td>
</tr>
<tr>
<td>1;2-1;3</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>1;3-1;4</td>
<td>England</td>
<td>1;5 month</td>
</tr>
<tr>
<td>1;4-1;5</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>1;5-1;8</td>
<td>England</td>
<td>3 months</td>
</tr>
<tr>
<td>1;8-1;9</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>1;9-2;4</td>
<td>England</td>
<td>6 months</td>
</tr>
<tr>
<td>2;4-2;5</td>
<td>New Zealand</td>
<td>1 month</td>
</tr>
<tr>
<td>2;5-3;0</td>
<td>England</td>
<td>7;5 months</td>
</tr>
<tr>
<td>3;0-3;0</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>3;0-3;4</td>
<td>England</td>
<td>4;5 months</td>
</tr>
<tr>
<td>3;4-3;5</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>3;5-3;11</td>
<td>England</td>
<td>6;5 months</td>
</tr>
<tr>
<td>3;11-4;0</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>4;0-4;3</td>
<td>England</td>
<td>3 months</td>
</tr>
</tbody>
</table>

The communication patterns at home are typical of a bilingual family in which one of the parents is monolingual (the father does not speak Polish): English is spoken whenever the whole family is involved and between the parents. When the father is not present, the mother addresses the children in Polish, occasionally inserting English words (individual items that are easier for the child to pronounce, or proper names of English story characters). In conflict situations, misunderstandings, or whenever the children disobey the parents, the
mother tends to address the children in Polish, despite the father not being able to understand the content. The communication patterns between the siblings have changed with time. At the beginning of her own speech, Patrick’s older sister Ellie (2 years older) used Polish to address him. Growing up bilingually and going to the local school added English to the resource pool for communication with her little brother. However, in every day conversations Polish is still her first choice when addressing Patrick or her mother.

The first meeting with the family took place when Patrick was 2;1, but due to the family’s holiday arrangements, the recording sessions did not commence until the child was 2;5. The sessions continued for 22 months, until Patrick turned 4;3. As to the recorded speaker configurations, the majority of interactions involve Patrick and the investigator, the next most frequent are interactions between the child and his mother, then between the child and his sister, and then with the father. To obtain additional English data (the low frequency of conversations with the father and the sister are indicative of the small amount of English input in the recorded conversations), the investigator sometimes spoke English with the child who often chose to answer in Polish. Since on a few occasions, the mother and the investigator discussed the recording sessions using Polish, the child might have decided that the investigator was to be addressed in Polish, and was thus greatly surprised when one day she addressed him in English. After two or three switches from one language to the other, the child was no longer surprised that both languages were being used, nor was he uncomfortable to switch between the two languages himself. For two reasons, however, the recordings lack more examples of interactions between Patrick and his parents: the father would usually work late, and interactions involving him mostly happened at the dinner table; the mother chose not to interfere with the recording sessions, and during the investigator’s visits she would usually continue her everyday household chores. The few sessions where Patrick’s father joined the family for supper allowed the investigator to document a very natural bilingual environment in which Patrick and Ellie were (and still are) being raised.

Overall, the data consists of mainly Polish interactions between Patrick and the investigator, but there are also English interactions between Patrick and his father, or Patrick and the investigator. Table 11 below provides details of all the interaction arrangements. In a large number of recordings, especially from session 4 onwards, the mother was absent for either a half or most of the session. The (M) symbolizes a recording session when the mother was physically in the house, but did not participate. Usually, there would be some interaction between the mother and Patrick at the beginning of the visit, after which the investigator and the child would start playing, while the mother would leave the room to occupy herself with her own tasks (cooking, emailing, etc). She would then come from time to time to comment
on the activities, or join the investigator and the child(ren) for a short game. As explained earlier, there are no recordings of the father playing with the child.

### Table 11. Patrick’s linguistic environment at recording sessions

<table>
<thead>
<tr>
<th>Session</th>
<th>Duration</th>
<th>Adults present</th>
<th>Languages spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42&quot;00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>2</td>
<td>50&quot;00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>3</td>
<td>37&quot;00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>4</td>
<td>21&quot;00</td>
<td>Investigator (M)</td>
<td>Polish</td>
</tr>
<tr>
<td>5</td>
<td>61&quot;00</td>
<td>Investigator (M)</td>
<td>Polish, English</td>
</tr>
<tr>
<td>6</td>
<td>77&quot;00</td>
<td>Investigator (M)</td>
<td>Polish, English</td>
</tr>
<tr>
<td>7</td>
<td>32&quot;00</td>
<td>Mother, Father</td>
<td>Polish, English</td>
</tr>
<tr>
<td>8</td>
<td>87&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>9</td>
<td>43&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>10</td>
<td>23&quot;20</td>
<td>Mother</td>
<td>Polish, English</td>
</tr>
<tr>
<td>11</td>
<td>93&quot;00</td>
<td>Investigator (M)</td>
<td>Polish, English</td>
</tr>
<tr>
<td>12</td>
<td>48&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>13</td>
<td>40&quot;00</td>
<td>Investigator (M)</td>
<td>Polish, English</td>
</tr>
<tr>
<td>14</td>
<td>65&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>15</td>
<td>113&quot;00</td>
<td>Mother, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>16</td>
<td>45&quot;00</td>
<td>Investigator, (M)</td>
<td>Polish</td>
</tr>
<tr>
<td>17</td>
<td>87&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>18</td>
<td>118&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>19</td>
<td>101&quot;00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>20</td>
<td>93&quot;00</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>21</td>
<td>65&quot;00</td>
<td>Investigator, (M)</td>
<td>Polish</td>
</tr>
</tbody>
</table>

### 3.2.2 Family 2: Jerzy

Jerzy was born on 9 September 2004 in Poland and was exposed to both languages regularly from the day of birth onwards. At the time of Jerzy’s birth both parents were 30 years old. The family lives in London and throughout the entire recording period it remained a three-member family. The mother, Maria, was born and educated in Poland. She obtained her M.A. degree in Italian at the Jagiellonian University in Kraków and worked as a teacher before settling down in the UK in February 2005. Since Jerzy’s birth, she has not been employed outside the home. The father, Nathaniel, was born and educated in the UK. He received his B.A. degree from the School of Slavonic and East European Studies at University College London and for a few years continued his education in Poland, where he mastered Polish. At the time of data collection, he worked as an IT commercial manager. After the birth, Jerzy was wrongly diagnosed with a skin disease called epidermolysis bullosa, which forced the family “to live” in the hospital for an additional month. The boy
was regularly exposed to both languages due to the constant presence of his English and Polish speaking family. In the second month, Jerzy was discharged from hospital and his exposure to both languages continued thanks to his father’s presence in Poland. After the father’s return to England, the child and the mother stayed in Poland for further six weeks, during which the child was addressed in Polish. In total, Jerzy spent the first three a half months of his life in Poland. His exposure to English increased on his return to England, but thanks to his mother and the maternal grandmother his exposure to Polish continued as well. Trips to Poland continued regularly approximately every two to four months, but on most occasions, the child would travel with the mother only. The father visited Poland for a week at Easter 2007, but he usually stayed in England due to work commitments or the house repairs, which dragged mercilessly, affecting the family’s flexibility and availability. The entire history of Jerzy’s travels between 0-3;11 is listed in Table 12.

In this family, both parents speak each other’s languages. Each of the parents addresses the child in their native language, thus following the One-Parent-One-Language method. Both Polish and English is used in conversations between the parents, and when the discussion involves the Polish grandmother, or when the topic is related to Poland and the Polish culture, the main language of interaction is Polish. From the recordings made by the mother it also transpires that at dinner table the parents would switch from English to Polish while discussing various topics, but whenever it was necessary for either of them to discipline Jerzy, they would each address the boy in their own language. Individual English words coming from the mother or Polish words used by the father can be treated as natural elements of communication of a bilingual home. As far as the child is concerned, Jerzy always addressed his parents in their native languages, making every effort to follow the “one language with one parent” rule. He would never use English to address his Polish grandmother, and in interactions with his English family, he would use English only. Patterns of communication with the investigator have changed over time. Initially the child did not seem to mind being addressed in both English and Polish, and until the approximate age of 3;0, frequent switching did not pose a problem in smooth communication. After that age, Jerzy chose Polish as the main tool for communication with the investigator, and preferred the languages to stay separate to the extent that books, which his father read to him in English, he would read to the mother and the investigator also in English, even when the language of the whole session was Polish. The gradual increase of Jerzy’s linguistic awareness dictated the rules for the interactions between him and the other speakers. As is clear from Table 13 below, 55% of the sessions were entirely in Polish, and approximately 43% were in both languages. There could have been a few sentences or phrases/words in
English during the Polish-based sessions, but they were too few to qualify English to the language spoken actively to the child during that session.

Table 12. Jerzy's places of residence

<table>
<thead>
<tr>
<th>AGE</th>
<th>COUNTRY</th>
<th>APPROX. DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3;0</td>
<td>Poland</td>
<td>3 months</td>
</tr>
<tr>
<td>0;3-0;4</td>
<td>England</td>
<td>1 month</td>
</tr>
<tr>
<td>0;4-0;5</td>
<td>Poland</td>
<td>1 month</td>
</tr>
<tr>
<td>0;5-0;6</td>
<td>England</td>
<td>6 weeks</td>
</tr>
<tr>
<td>0;6-0;7</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>0;7-0;9</td>
<td>England</td>
<td>2 months</td>
</tr>
<tr>
<td>0;9-0;9</td>
<td>Poland</td>
<td>2 weeks</td>
</tr>
<tr>
<td>0;9-0;9</td>
<td>England</td>
<td>3 weeks</td>
</tr>
<tr>
<td>0;9-0;10</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>0;10-1;0</td>
<td>England</td>
<td>2 months</td>
</tr>
<tr>
<td>1;0-1;1</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>1;1-1;3</td>
<td>England</td>
<td>2 months</td>
</tr>
<tr>
<td>1;3-1;5</td>
<td>Poland</td>
<td>2 months</td>
</tr>
<tr>
<td>1;5-1;8</td>
<td>England</td>
<td>3 months</td>
</tr>
<tr>
<td>1;8-1;10</td>
<td>Poland</td>
<td>2 months</td>
</tr>
<tr>
<td>1;10-2;1</td>
<td>England</td>
<td>3 months</td>
</tr>
<tr>
<td>2;1-2;2</td>
<td>Poland</td>
<td>6 weeks</td>
</tr>
<tr>
<td>2;2-2;6</td>
<td>England</td>
<td>4 months</td>
</tr>
<tr>
<td>2;6-2;8</td>
<td>Poland</td>
<td>6 weeks</td>
</tr>
<tr>
<td>2;8-3;0</td>
<td>England</td>
<td>4 months</td>
</tr>
<tr>
<td>3;0-3;2</td>
<td>Poland</td>
<td>2 months</td>
</tr>
<tr>
<td>3;2-3;3</td>
<td>England</td>
<td>5 weeks</td>
</tr>
<tr>
<td>3;3-3;3</td>
<td>Poland</td>
<td>2 weeks</td>
</tr>
<tr>
<td>3;3-2;5</td>
<td>England</td>
<td>2 months</td>
</tr>
<tr>
<td>3;5-3;6</td>
<td>Poland</td>
<td>3 weeks</td>
</tr>
<tr>
<td>3;6-3;8</td>
<td>England</td>
<td>6 weeks</td>
</tr>
<tr>
<td>3;8-3;10</td>
<td>Poland</td>
<td>2 months</td>
</tr>
<tr>
<td>3;10-3;11</td>
<td>England</td>
<td>7 weeks</td>
</tr>
</tbody>
</table>

The investigator first visited the family in August 2006, just before Jerzy turned two, and a month before the recordings commenced. This initial contact allowed all speakers to get used to each other and become more confident around each other. Similarly to Patrick’s family, a variety of interactions has been recorded. Yet, contrary to sessions with Patrick, many recordings with Jerzy involve him and his mother. It partly results from the fact that the mother usually tried to provide the most natural linguistic environment by being involved in the games, and partly due to the frequent trips to Poland. In order to continue the sessions, the mother simply had to take the recorder with her, and as a result, those sessions include the mother, and often the grandmother. Conversations between the mother and the child are
most numerous. Next are conversations between the child and the investigator, and then with
the father and grandmother. Recorded interactions between Jerzy and his English
grandparents are very scarce. It is clear from Table 13 that the sessions are not comparable in
terms of length: those recorded by the mother tended to last 15-30 minutes and were
arranged on different days. The sessions recorded by the investigator were longer and often
bilingual. During a session with the investigator, the routine would involve a brief chat with
the mother about recent changes in the child’s speech, then a spontaneous play or a chat with
Jerzy, after which the session proper would begin. Jerzy’s mother has been invaluable in
“navigating” a large number of sessions. Her intuition and knowledge has always been
extremely helpful, and her involvement has provided the investigator with useful knowledge
about the child’s character and development. A wide array of activities has been recorded,
starting with reading children’s books, car magazines and animal encyclopaedias, through
painting, acting-out, describing pictures, to pretend phone-calls, shopping or eating out. At
times, the mother would only observe the playing, commenting on it from time to time or
asking the child questions. At other times she would be actively involved, especially in
reading books or showing the family pictures, which Jerzy enjoyed describing in every
detail. Any interactions between the father and the child have been recorded by the mother
(hence the lack of video recordings for those sessions). It must be stressed that both Jerzy’s
parents have shown a great sensitivity to the needs of this project by providing both
maximum flexibility and a great understanding of the nature of fieldwork. Recordings done
by the mother include natural interactions between the family members, rather than arranged
sessions during which parents would ask the child guided questions. They are varied, active,
and most importantly, they encourage the child to talk. Overall, the 31 Polish-based sessions
sum up to nearly 19 hours, while the Polish-English sessions amount to approximately 14
hours. Determining the exact amount of English and Polish during the bilingual sessions
would require manual counting of the words/utterances produced by the speakers. The
unabridged data adds up to 45 hours of recordings.

Table 13. Jerzy’s linguistic environment at recording sessions

<table>
<thead>
<tr>
<th>Session</th>
<th>Duration</th>
<th>Adults present</th>
<th>Languages spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20''00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>2</td>
<td>35''00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>3</td>
<td>45''00</td>
<td>Mother, Father</td>
<td>Polish, English</td>
</tr>
<tr>
<td>4</td>
<td>60''00</td>
<td>Mother, Grandmother, Father</td>
<td>Polish, English</td>
</tr>
<tr>
<td>5</td>
<td>24''25</td>
<td>Mother, Grandmother</td>
<td>Polish</td>
</tr>
<tr>
<td>6</td>
<td>26''00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>7</td>
<td>90''00</td>
<td>Mother, Grandparents</td>
<td>Polish, English</td>
</tr>
<tr>
<td>8</td>
<td>67''25</td>
<td>Mother, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>9</td>
<td>27''00</td>
<td>Mother</td>
<td>Polish</td>
</tr>
<tr>
<td>No.</td>
<td>Age</td>
<td>Relationship</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>43'45</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15'00</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>23'20</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>27'00</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>58'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>15'50</td>
<td>Mother, Grandmother, Father</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>24'15</td>
<td>Mother, Grandmother, Father</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>22'00</td>
<td>Mother, Grandmother, Father</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>21'00</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>17'30</td>
<td>Mother, Grandmother, Father</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>24'30</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>18'30</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>20'20</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>27'35</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>47'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>58'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>56'45</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>30'05</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>24'10</td>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>24'00</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>61'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>21'25</td>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>33'35</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>9'50</td>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>93'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>14'00</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>11'20</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>20'00</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>22'10</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>31'05</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>14'40</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>20'00</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>17'45</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>21'45</td>
<td>Mother, Grandmother</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>72'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>103'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>95'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>96'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>75'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>96'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>13'20</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>17'10</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>28'50</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>82'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>44'20</td>
<td>Mother, Father</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>33'00</td>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>60'00</td>
<td>Mother, Investigator</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>35'00</td>
<td>Father</td>
<td></td>
</tr>
</tbody>
</table>
3.2.3 Family 3: Hania

The third family investigated here also lives in London. On the day of Hania's birth, her mother was 33, and her father was 51. The mother, Matylda, was born and brought up in Poland, where she graduated to be an artist (sculptor), and she has worked outside home since the graduation. After settling down in the UK in 2002 and giving birth to Hania, Matylda tried to set up her own art business. It involved working outside home and leaving Hania in the care of her grandmother, aunt or a baby-sitter. Matylda was not in paid employment when Hania turned 2;6, and remained at home due to pregnancy and the birth of her second child. The father, Keith, was born and educated in England, where he has been employed as a carpenter. Since he has two sons from his first marriage, Hania is his third child. The two half-brothers are approximately 13 and 15 years older than Hania.

Hania was born in the UK on 6 July 2004. Her exposure to Polish and English began on the day she was born, and has continued thenceforth. With Hania's birth, the family employed the One-Parent-One-Language strategy. In the first three weeks of life the girl was exposed to Polish and English in equal amounts, and after that her Polish grandmother visited the family in order to support the mother, thus increasing Hania's daily exposure to Polish. Although she received English input from her father and conversations between the parents, Polish remained dominant for at least three years. The father would spend 5 or even 6 days at work, while the child was looked after either by the mother, grandmother or a Polish-speaking nanny. When Hania turned one, Keith began to spend more time playing with her, and from then on, they would spend two hours nearly every evening playing and chatting in English. Similarly to Jerzy and Patrick, an important element of Hania's bilingual upbringing were trips to Poland and visits paid by the Polish relatives. Let us look at Hania's exposure patterns in a chronological order. Between the age of 2-7 months, a Polish nanny was employed to look after Hania for approximately 10 hours every day while the mother worked. The Polish grandmother provided additional input in the first 2 months of the child's life as well as between the age of 7-9 months. Hania was 5 months old when she first visited Poland, whereas her second trip took place before her first birthday. Both trips lasted two weeks. Between the child's age of 9-24 months Matylda was not in paid employment, which allowed her to stay at home, but this arrangement changed soon after the girl turned 2;0. With Matylda's new job, there was a need for another helping hand. First, Hania was in the care of her Polish cousin, and after a month her Polish aunt stayed with the family until the girl turned 2;5. After the aunt's departure, Hania was with her mother again, as Matylda's job did not prove to be worth investing the time, which also allowed the recording sessions to commence. At the age 2;6 Hania visited Poland for the third time, again for two weeks. Meeting her cousins and other relatives certainly boosted her Polish. Her next trip to
Poland was a six-week holiday between 3;0-3;1, which had an even bigger impact on the child’s Polish. Interestingly, English lagged behind Polish until the age of 3;0, and the gap seemed even bigger after the child’s long stay in Poland. Nevertheless, soon after she had joined the nursery at the age of 3;2, there was a noticeable growth in the complexity of the sentence structure as well as in the English vocabulary. Evidently, English caught up with Polish very quickly. At the nursery, Hania spent two and a half hours a day until she began preschool at the age of 4;0. In time, the exposure patterns to both languages became stable and regular. Table 14 lists Hania’s places of residence.

Compared to Jerzy’s, Hania’s visits to Poland seem infrequent and rather short. However, the Polish input she received whilst living in the UK can be regarded as equally effective as that which Jerzy received when travelling to Poland every 3-4 months. All carers involved in looking after Hania were Polish and spoke standard Polish. Additional resources for Polish (books and DVDs) were used by the mother, who tried to make sure that her child was going to speak Polish without delays and problems. There has never been a period of time when she would refuse to be read a book in either language, which shows that her growth in both languages, albeit unbalanced, has been comfortable for her. Overall, between the age 0-4;1 Hania visited Poland 4 times for a total of 3 months, and the remaining time she spent living in the UK with her parents and carers.

The investigator met the family twice before data collection process began (when Hania was 2;2 and 2;3). Throughout this initial acquaintance, the investigator used Polish with Hania and her mother, which prompted the child to link the investigator with Polish. As can be seen below (Table 15), the number of bilingual sessions equals that of sessions involving only Polish. Determining the exact amount of Polish and English during the bilingual meetings would be possible only after counting all the utterances manually. Nonetheless, it is clear without such count that English data is less dense than Polish data. The investigator tried to prompt the child to use English with her and succeeded in doing so between the ages 2;4-3;2. After the child started attending the nursery five times a week, her linguistic awareness and the need to obey the “one-person-one-language” rule demanded that interactions with the investigator took place in Polish. As a result, between the age 3;2-3;9 the child would rarely pick up and continue a conversation with the investigator in English. From time to time, the investigator used English with the child and prompted her to speak English back, but she was far less willing to do so once she joined the nursery. Interestingly, at the age 4;1 a change occurred. The child’s determination to use Polish with the investigator decreased, allowing a large chunk of some sessions to take place in English. Table 15 shows the linguistic environment at the recording sessions. Initially, the mother was present more often, but over time the dynamics of the meetings changed. As the majority of
sessions took place before noon, they usually began with conversations over breakfast and
the mother would always be involved. After breakfast, the investigator and the child would
begin playing, and the mother would be around in the house doing other things, and from
time to time stopping for a chat. The mother usually preferred to leave the investigator and
the child to play while she tried to deal with job hunting and domestic chores. On a few
occasions, the investigator left the recorder for the mother to use. In each of those sessions,
the mother-child interactions and the father-child interactions are focused on playing and
actual language production. They hugely contributed to the quantity/quality of the data.

Table 14. Hania’s places of residence

<table>
<thead>
<tr>
<th>AGE</th>
<th>COUNTRY</th>
<th>APPROX. DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-0;5</td>
<td>England</td>
<td>5 months</td>
</tr>
<tr>
<td>0;5-0;6</td>
<td>Poland</td>
<td>2 weeks</td>
</tr>
<tr>
<td>0;6-0;11</td>
<td>England</td>
<td>6 months</td>
</tr>
<tr>
<td>0;11-1;0</td>
<td>Poland</td>
<td>2 weeks</td>
</tr>
<tr>
<td>1;0-2;6</td>
<td>England</td>
<td>18 months</td>
</tr>
<tr>
<td>2;6-2;6</td>
<td>Poland</td>
<td>2 weeks</td>
</tr>
<tr>
<td>2;6-3;0</td>
<td>England</td>
<td>5 months</td>
</tr>
<tr>
<td>3;0-3;1</td>
<td>Poland</td>
<td>6 weeks</td>
</tr>
<tr>
<td>3;1-4;1</td>
<td>England</td>
<td>12 months</td>
</tr>
</tbody>
</table>

A few more details need to be mentioned with regard to the family’s everyday
communication patterns outside the recording sessions. Since the father does not speak any
Polish, the parents always use English to address each other, and any conversations
involving both parents and the child are always carried out in English. Polish is the main
communication tool between the mother and the child apart from the instances involving the
whole family when the mother addresses the girl in English to avoid excluding the father
from the interaction. The grandmother would address the child in Polish, and the rare
interactions between the grandmother and the father would be interpreted by the mother.
Any other Polish relatives unable to speak English would communicate with the father via
the mother.

Table 15. Hania’s linguistic environment at recording sessions

<table>
<thead>
<tr>
<th>Session</th>
<th>Duration</th>
<th>Adults present</th>
<th>Languages spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38”00</td>
<td>Mother, Investigator, Baby-sitter</td>
<td>Polish</td>
</tr>
<tr>
<td>2</td>
<td>50”25</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>3</td>
<td>15”30</td>
<td>Mother, Father, Investigator</td>
<td>Polish, English</td>
</tr>
<tr>
<td>4</td>
<td>60”00</td>
<td>Mother, Investigator</td>
<td>Polish</td>
</tr>
<tr>
<td>5</td>
<td>36”00</td>
<td>Mother, Father</td>
<td>Polish, English</td>
</tr>
</tbody>
</table>
In sum, the total number of unabridged recordings with Hania adds up to 53 hours. A large number of recorded interactions involve the investigator and the child. The next most frequent are mother-child interactions, followed by those involving the father and the child. The few sessions that include the child interacting with a baby-sitter, a relative, a visitor or her Polish grandmother add some richness and variety to the overall pool of interactions.
3.3 MLU

The year 1973 witnessed the appearance of Brown’s famous study on Adam, Sarah and Eve (see §3.1) and the well-known language growth index (MLU), which Brown used for calculating the average length of an utterance in the children’s speech. In his pioneering work on the development of the “fourteen English morphemes” by the three children, Brown (1973) suggested MLU as a simple index of grammatical growth on the basis that as children begin to acquire grammar, they produce utterances that include more and more grammatical elements, such as plural or case markers, articles etc. Brown’s reliance on the concept of “a morpheme” led him to a list of 14 linguistic elements, which mark the child’s gradual linguistic development. They are: prepositions (in, on), articles (a, the), plural -s, possessive ‘s, progressive -ing, past tense -ed, 3rd person verb -s, irregular past tense, irregular 3rd person verb, and various forms of be (uncontractible and contractible copula, uncontractible and contractible auxiliary). It was crucial to identify those “morphemes” in child’s utterances to track the progress in her developing grammar. However, MLU can be measured not only in morphemes (MLUm), but also in words (MLUw) and clauses (MCU). Calculating MLU in morphemes enabled Brown to detect new linguistic elements to a larger extent than MLUw would have. By treating, e.g., the past tense marker -ed in played or the genitive -s in mummy’s as an autonomous element, MLUm acknowledged the fact that the child has learned a new morphological feature, and added it to the score as two separate morphemes. MLUw, on the other hand, would count both play and played, or mummy and mummy’s as one morpheme. Interestingly, in their recent comparative study between MLUw with MLUm in the speech of English-speaking children, Parker and Kent (2005: 265) report that both indices are almost perfectly correlated, i.e. “MLUw can be used as effectively as MLUm as a measurement of a child’s gross language development”. Over the years, MLUm has become the standard index for research conducted on English data. However, researchers dealing with highly inflected languages have had doubts about the reliability of MLUm, mostly due to linguistic differences between languages and differences between the definitions of “utterance” and “morpheme”. Unsurprisingly, “its applicability to the study of other languages has been under heated debate” (Schnell de Acedo, 1994: 250), and even Brown admitted that there is some “difficulty in adapting the rules of calculation invented for English” (1973: 68). It must be remembered, however, that “most of the measures that are commonly used to analyse spontaneous language production were originally developed for English-speaking subjects” (Schnell de Acedo, 1994: 250), and it is challenging to employ any morpheme counting index to a language other than English. These initial concerns as to the applicability of MLUm, however, have been recently overshadowed by the criticism of
the concept of "a morpheme" as a linguistic phenomenon in general. The following sections discuss this problem in more detail.

Following the tradition of Hanrican structuralist linguistics established by Bloomfield (1933), The Dictionary of Language Teaching and Applied Linguistics defines morpheme as "the smallest meaningful unit in a language" and continues:

A morpheme cannot be divided without altering or destroying its meaning. For example, the English word *kind* is a morpheme. If the *d* is removed, it changes to *kin*, which has a different meaning. Some words consist of one morpheme...For example, the English word *unkindness* consists of three morphemes: the stem *kind*, the negative prefix *un-*, and the noun-forming suffix *-ness*. (Richards, Platt and Platt, 1985: 236)

Hockett (1947) follows this the same path in defining morpheme, while O’Grady expands it to “the smallest unit of language that carries information about meaning or function” (1997). Unfortunately, such definitions are open to objections due to their very general nature. Firstly, they concentrate on the spoken or written language, or make reference to semantics, but fail to refer to the sound of spoken language. Secondly, and more importantly, morphemes segmentable in one language are not as easily identifiable in another, e.g., English derivational morpheme *un* or the plural grammatical morpheme *-s* have clear meaning assigned to them, but there is no segment in the German *Mütter* ‘mothers’ (vs. *Mutter* ‘mother’) that can be assigned the meaning ‘plural’. As Haspelmath explains, it is “a clear-cut example of morphological structure in that a recurrent meaning (‘plural’) corresponds to a recurrent aspect of form (the front vowel), but the plural word-forms cannot be segmented” (2002: 18). The third important aspect is the relationship between morphemes and allomorphs. Let us consider the English plural morpheme *-s*, which is realised by three different allomorphs: /-s/ *cups*, /-z/ *tables* /-iz/ *houses*. The situation becomes more problematic when we take into consideration more challenging examples, such as two English past participle suffixes: *-en* as in *forgiven* and *-ed* as in *stopped*, and try to decide whether they are suppletive allomorphs of the same morpheme, or two completely different morphemes. Haspelmath (2002: 31) points out that if those two endings are manifestations of the same morpheme, then the traditional definition of a morpheme becomes woolly. The concept of a morpheme in such case would need to be both, a concrete minimal morphological constituent with a function, such as the English plural *-s*, as well as an abstract notion, a sort of "metaphor", defined as “the set of alternating morphs that have the same meaning and occur in complementary distribution” (Haspelmath, 2002: 31). The confusion does not end here, since the term morpheme also functions as an “umbrella” term for such concepts as affix, stem, or root, which have clear-cut definitions of their own and do
not need an ambiguous replacement. Let us now consider potential implications resulting from the above. Any in-depth morphological analysis of smallest meaningful elements requires access to the parts of language that may manifest themselves differently in different languages, thus due to its ambiguous character the term morpheme cannot be used as the basis for such an analysis, especially in a comparative study of two morpho-syntactically different languages, such as Polish and English. In Brown’s view, morpheme is a countable element of language, yet in the metaphorical sense it can only be an abstraction. As a result, linguists have rejected the notion of a morpheme altogether, arguing that the abstract nature of a morpheme makes it superfluous and fundamentally flawed, and as such it can be either treated as a metaphor only, replaced by a different conception or totally avoided (Spencer, 1991, Anderson, 1992). Since morpheme can no longer be unambiguously defined, measuring the mean length of a child’s utterance by counting the number of morphemes is too unreliable for the purposes of this study.

Identifying the set of morpho-syntactic and morpho-semantic properties acquired by a child is a necessary step towards an adequate description of the child’s linguistic development on the morphosyntactic level. Such analysis can be difficult not only from a comparative perspective in which comparing like with the like is a necessary requirement, but also when measuring the growth of the structural complexity within one language. In modern linguistic theory, the concept whose definition has been attempted in the earlier paragraphs is a feature. A given feature can be manifested in a language without having a separable morphological representative. Languages with fusional morphology, such as Polish, tend to collapse many morphemes in a way that is difficult or impossible to segment. Consider the example below:

(36) Spacerując przeglądałyśmy się ludzi-om na ulicy.
while.walking we.were.observing people-PL.DAT in the street.
“While walking, we were observing people in the streets”

The affix -om attached to the noun ludzie ‘people’ results in ludziom ‘people-PL.DAT’, which demonstrates how number and case can be fused into one affix; none of those features can be identified separately. It is, therefore, more adequate to say that a certain affix in a fusional language can denote a cluster of features, and changing any of these features requires replacing the affix with something else. Linguistic description starts with the identification of a set of morphosyntactic features and accounts for the way the features are realised in the language’s morphology. Consequently, due to the criticism expressed in the paragraphs above, MLU in morphemes cannot be used for calculating the growth of structural complexity in Polish, since such complexity cannot be traced by simply choosing 14
important suffixes and counting them in the child’s speech. Similar concerns can be expressed with regard to MLU in words. Since MLU is a general index of grammatical development, it does not reflect the increase of morphological or semantic sophistication in the child’s language. Nonetheless, counting words using CLAN can give us an idea of how the child’s lexical resources are changing with age, how that growth can be correlated to the input, and much more.

Before we proceed to discussing the database, it needs to be clarified what is meant by saying “calculate MLU in words”. The Encyclopedia of language and linguistics (Brown, Asher and Simpson, 2006) lists a few ways of understanding this concept. Firstly, there is a phonological word, which may be a lexical item combined out of two (or more) syntactically independent elements (e.g., in I’ll come, the item I’ll is seen as one phonological word). Outside of the phonology both I and will have their own representation in the syntax, and are therefore treated as two separate words. Secondly, there is a lexeme which has been defined as “an entry in the speaker’s mental lexicon” (Julien, 1993: 618), or as “a dictionary word” (Haspelmath, 2002: 13). Each lexeme entered in the dictionary subsumes inflected forms that this lexeme can take. Thus, a lexeme pies ‘dog(My)’ is listed only to point the reader to other forms of this word, such as ps-a ‘dog-GEN,SG’, ps-y ‘dog-NOM,PL’, etc. These other forms are called grammatical words (Julien, 1993: 618), or word-forms (Haspelmath, 2002: 13), and are used in writing or speech. MLUw calculates the number of word-forms used by a child, i.e. every word produced by a child is counted as separate, even if there are a few word-forms belonging to one lexeme in the same utterance. Based on the database compiled for this project, it has been observed that counting word-forms allows us to track the child’s lexical development, but it does not provide any information about the child’s morphosyntactic sophistication. At the stage of rich spontaneous production, Hania’s utterances tended to be long as they involved repetitions. The word count for the longest utterance at the age of 4;0 is 29, whereas the longest utterance at the age of 4;1 is only 23. Compare below:

Age 4;0 CPL: nie, bo [bo -- ja --] ja nie lubię tak [tak -- tak -- tak] mówić do many, bo [bo ja --] bo ja jestem smutna, bo ja [nie ce --] nie chcę- są mówić do many, bo ja [ja -- ja -- ja] nie lubię, żeby było [ma -- ma --] bardzo -- baldzo -- baldzo --] dużo ludzi=ludzi@.

‘no, because [because -- I --] I don’t like so [so -- so -- so --] speak to mum, [because I --] because I am sad, because I [don’t want --] don’t want to talk to mum, because I [I -- I -- I --] don’t like when there are [ma -- ma --] very [very -- very --] many people.’

Age 4;1 CPL: ale [ale --] tak byłem zimny, to [to --] jak śnieg padał tam [tam --] w Polsce [to -- ja s-- to ja -- to ja -- to mia --] miałam [nie -- nie rękawiczek=ląkawicki@ to [to --] spadam i pan doktor=doktor to [....] tam przyszeli. [#CLS]

‘but [but --] when I was cold then [then --] when it was snowing there [there --] in Poland then [-- then I -- then I -- then I -- ha --] I had [no --] no gloves then [then --] I fell and the doctor here [....] there came.’
Hania’s five longest utterances at 4;0 contained 85 word-forms, while at 4;1 they contained 81 word-forms. An examination of the ratio of words over the five longest utterances at the two age stages might lead to a conclusion that Hania’s lexical productivity was greater at 4;0 than at 4;1, which is a reasonable suggestion as long as we realise that to a certain degree Hania’s MLUw is increased by a certain type of repeated material. Importantly, MLUw shows us how the child’s lexicon expands, but it does not reflect the grammatical complexity of the child’s language. Hania’s utterances at 4;0 might be slightly longer than those at 4;1, but a more detailed examination shows that the older the child gets the more complex her utterances become. The longest utterance at 4;0 contains a recursive element ‘because’, which introduced new information to the utterance. The longest utterance at 4;1 is a complex temporal clause relating to some past events. As an additional illustration to the general limitation of MLUw, consider here an example from North Sami (North Saami): beatnagiiddisetguin ‘with their (own) dogs’ (Julien, 1993: 618). MLUw would count it as one word, while the complexity of this phrase requires much greater attention and appreciation. A child who learned to combine correct items into one word of such sophistication would need to be regarded as more proficient in her language than a child who produced one word of a much simpler nature, e.g. beana ‘dog’ (Julien, 1993: 618). Both children would score one point in the MLUw index, which is undoubtedly an example of misrepresentation of the actual linguistic development of one of the children.

An important issue requires clarification here. As advised by MacWhinney, the creator of CLAN, decisions as the inclusion or exclusion of the child’s imitation and self-repetitions in/from the MLU count and other CLAN procedures can be made by the researcher. In this project, repeated material produced by the participants has been excluded from CLAN analyses to prevent artificially increased scores (square brackets were used in the coding process). Any analyses including imitation have been be reported as such. Importantly, whenever a child finished an utterance, and then decided to repeat it after the interlocutor’s turn, the child’s utterance was not marked as repetition.

3.4 Data collection

Quantitative research investigates quantitative properties of various phenomena with the use of mathematical measuring tools, such as statistics. Qualitative research deals with human behaviour in a context and attempts to investigate and understand it via in-depth analysis. Qualitative research is naturalistic and interpretative, and the sources of data are field notes, recordings, conversations and observation. The study reported here is qualitative in nature. To study language and its development, “researchers often observe, record and analyse what
children naturally and spontaneously say. Such naturalistic observation has the undeniable appeal of ecological validity and is often considered objective and reliable” (Bornstein, 2002: 687-8). Yet, observations are per se constrained by the presence of an observer. This phenomenon has been defined by Labov as the Observer’s Paradox: “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation” (Labov, 1970: 3). In other words, the observation of an event cannot be fully objective and natural due to the presence of the observer, whereas removing the observer from the scene makes the observation impossible, and invalidates the naturalistic method of data collection. Fortunately for this study, although it is not free from the “observer’s paradox”, and despite the children feeling some initial reservations, they soon became open, natural and very spontaneous, quickly making (some in a quite a “possessive” way) the investigator/observer their partner in playing and chatting. The data investigated here comprise a detailed case study of three children: Patrick, Jerzy and Hania. As has been explained in §3.1, the recruitment criteria were strict for the data to be comparable, and a number of factors underlying the choice of participants has been taken into account: the children’s own linguistic balance in the two languages, the mothers’ availability for recording sessions, and their willingness to record the children on their own. Databases for Patrick, Jerzy and Hania were the richest and the most complete, whereas the parents’ willingness to record their interactions with their children made a significant difference to the final quality of the database of their own child. Due to the context-bound nature of fieldwork-based research such as the present study, a certain amount of unpredictability was inescapable. The investigator had to allow the probability of gaps in recording sessions during the data collection process resulting from: sudden and unplanned trips to Poland, parents changing jobs, children being ill, the house being redecorated, etc. The initial situation in which each family was found when entering the project usually altered over time due to some unforeseen factors. The same applies to the quality of the input. Any generalisations here might result in an unfair judgement. However, mere observation of everyday situations is sufficient to note that the stricter the mothers were about speaking the minority language (Polish) to the children, the more balanced the child’s language use seemed to be later on. Contrary to the three main subjects, two younger children, Malcolm and Mikolaj, were dominant in English at the time of the recruitment. Malcolm remained dominant in English until the last recording, whereas Mikolaj went through stages when Polish became more active, and then returned to being the minority language. As it happens, at the time of writing this paragraph, Mikolaj has become a balanced bilingual. Unfortunately, due to the project limitations, the recording sessions had to stop, and any changes, however useful or intriguing, needed to be
left out. A certain amount of unpredictability as well as luck is always present in research involving such young speakers.

In adopting the observational strategy, researchers are expected to make decisions about recording and analysis. Questions appear: when, where, for how long, how frequently, and with whom to record the child in order to be satisfied, and trust that the child’s language has been sampled appropriately. The overall plan for this study was to record spontaneous bilingual data as regularly as possible. Initially the aim was to collect the samples fortnightly, but due to the limitations of a one-person project, it soon regularized into every three weeks. Every now and then, the sessions were organised even less frequently for various reasons. The initial plan also aimed at collecting a balanced amount of data in each language. However, the absence of the English-speaking fathers at the sessions forced the investigator to compromise the data, and hope that the mothers would be willing to record samples of the children’s interactions in English when the fathers return from work. In bigger projects, this obstacle is usually overcome by inviting an under-graduate or a post-graduate student to join the investigator. In this project, such a student would use English only, so that the child could get used to two languages being used regularly at each session. The data would have been more balanced and transparent if the children were to produce equal amounts of speech in each language at each session. However, financial restrictions did not allow such a solution to be implemented here. Another challenge for a child language researcher comes from the fact that “no one situation typifies child language production or gives an indication of the child’s spontaneous verbal competence” (Bornstein, 2002: 688). In other words, the investigator must obtain a wide range of situations to collect a contextually rich and dense database. Every effort was made here to keep the sessions varied and interesting by providing ideas for playing or bringing new games and books for the children to explore. More information on naturalistic data sampling and useful guidelines to a methodologically more efficient data collection process can be found in Bornstein (2002) and Wells (1979).

We now turn to the recording equipment. For this project data collection was carried out in the children’s homes using the following tools:

- Edirol R-1 Solid State Digital Handheld Recorder
- Marantz PMD670 Professional Solid-State Recorder
- Canon Elura 100 camcorder
- MacArthur Communicative Development Inventory (CDI)

Due to the simplicity of use and the lightweight construction, the Edirol R-1 recorder and the Canon camcorder soon became the investigator’s main tools. The recorder was placed on the
floor, the table or a chair close to where the interaction was taking place, and it received little interest from the children, except of some infrequent occasion when the investigator was asked to explain what “the box” was there for and why it had so many buttons. There is no reason to believe that the interactions and the linguistic choices made by the speakers were influenced by the presence of either the recorder or the camcorder. Unfortunately, although at the time Edirol R-1 was the most recommended recorder, some valuable data was lost due to a fault, which would occasionally corrupt a file in the process of recording. Since approximately 1-2 files per 10 failed to record, the investigator made sure that the camcorder was recording everything that could have been of use. That way, if a file did not record on the Edirol R-1, the camcorder was there to make sure it was backed up on the tape. As mentioned in the preceding section, some families travelled to Poland frequently and to provide continuity in sampling, the mothers were asked to take the Edirol R-1 recorder with them. A problem occurred during periods such as Christmas, Easter, or summer holidays when trips to Poland would overlap. Gaps in recordings had to be allowed and some planning had to be applied as to which family was taking the recorder and when. All the files were recorded in WAV format ensuring high quality of the recorded data. Video-recording the sessions allowed for the context of the interactions to be present and clear. There is no video context for recordings made by the mothers, but to obtain an even fuller picture of the children’s speech development, the MacArthur Communicative Development Inventory (CDI) was used. A CDI is a list of words and phrases suggested as a basis for a particular stage of the linguistic development. The mothers were asked to fill the inventories, i.e. tick off all the words and phrases available on the list, plus add any English or Polish words that the child used. Two types of CDI were used for this study. Stage I CDI was focused on words and gestures, and it was given to the mothers every month between the first meeting and the moment their child began using morphological clues more actively when building their sentences. Stage II CDI was focused on words and sentences, and it was filled in on a monthly basis until the children turned three. In this project, Smoczyńska’s adaptation of the Inventories for Polish has been used. Additionally to the recorded data, one of the mothers (Maria) made the effort to keep an “early stage” diary, in which she noted down Polish and English words and phrases produced by Jerzy before the child’s utterances became longer and more complex making it impossible to keep the diary up to date.

In child-language-based studies such as this one, data collected from spontaneous interactions in naturally occurring situations was ideal. Yet, visiting the subjects and recording the interactions can be compared to the pleasure of buying oneself a musical instrument if planning to learn to play it. After the purchase, a more demanding stage begins...
- the practising. Similarly, simply collecting the data is only the beginning of a much longer journey. Language fieldwork researchers believe that a rule of thumb is that 1/3 of the project time will be spent in fieldwork, 1/3 in analysis, and the final 1/3 in writing up the work. Feagin (1996) clarifies that “though far from scientific, this rule provides an effective reminder of the point that time required for analysis and writing increases in a ratio of about 2:1 for each hour of data elicitation.” It is important to bear this rule in mind if we plan to collect raw data, and remember that the question of how much data we plan to collect is usually guided by the topic of our investigation. In this project, tracking down the development of grammatical gender from the earliest morphological stage to the signs of a learning strategy is certainly not a matter of a year. The investigator expected the recording sessions to continue for at least a year and a half, and with the ratio of 2:1, the analysis and the writing up were expected to take at least up to additional two to three years.

3.5 Corpora

Today’s child language researcher planning a session with a subject automatically reaches out for a recorder without even realising how much has changed in the area of child language research since the portable tape (and then digital) recorder was invented. Data collection was made easier and far more reliable. The method of analyzing, however, still lagged behind. MacWhinney (1994: 409) gives an interesting illustration of the “past”:

The new generation of child language researchers may tend to take precise computational tools for granted, but those of us who have spent years labouring with hand-written transcripts understand how far we have come from the days of diary notebooks, hand-compiled concordances, and blurred mimeographed copies. With the new tools [...] child language researchers no longer need to spend hours poring over transcripts looking for a single use of a word. We no longer need to mark tallies of word occurrences in the margins of our printed transcripts and then turn through our notebooks, page by page, adding up these tallies by hand, only to realise after hours of work that we have been ignoring some crucial dimension and that the whole analysis has to be started again from scratch.

It has been a successful and a productive, albeit a long journey - from hand-written diaries, through typewriters, mimeographed copies and scribbles on the margins to digitally recorded and automatically coded data. The results of this quest have a great significance today. By designing an automatic coding system and computational analytical tools, linguists and psycholinguists such as Brian MacWhinney, Ursula Bellugi, Courtney Cazden or Colin Frazer made it possible for thousands of researchers to “conceive of and carry out analyses more complex and subtle than could be dreamt of with Mongol pencils in hand” (Brown, 1994: ix). Such step was a logical follow-up to the invention of a portable recorder, and was
made possible by another electronic revolution, namely the emergence of affordable personal computers. The first library of child language data was created electronically, simplifying the storage, exchange and analysis. However, for the data to be comparable and trustworthy, a standardized transcription format needed to be invented and applied to any data that was to be included in the child language data library. Furthermore, the coding and the analysis needed to be automated, so that it could be consistent, error-free and verifiable. All those goals have been achieved by the Child Language Data Exchange System (CHILDES) created by Brian MacWhinney in the 1980s. The reasons for developing a computerized system for language data exchange are immediately obvious to anyone who has ever worked with naturalistic data and has tried to produce and analyse transcripts. The CHILDES system has developed three integrated, yet separate, tools:

1. CHAT - Codes for Human Analysis and Transcripts, i.e. transcription and coding format (§3.5.1)
2. CLAN - Computerized Language Analysis (§3.5.2)
3. the child language database (§3.2)

MacWhinney (2000) calls this three-tool system a “three-legged stool” and explains the interrelations between the three tools in the following way:

The transcripts in the database have all been put into the CHAT transcription system. The program is designed to make full use of the CHAT format to facilitate a wide variety of searches and analyses. Many research groups are now using the CHILDES programs to enter new data sets. Eventually, these new data sets will be available to other researchers as a part of the growing CHILDES database. In this way, CHAT, CLAN, and the database function as a coarticulated set of complementary tools.

The first step to approach raw naturalistic data is to transcribe/computerize it. Let us have a closer look at the transcribing rules employed in this project.

3.5.1 CHAT

The investigator followed the updated CHAT manual available on CHILDES website, which provides a “set of computational tools designed to increase the reliability of transcriptions, automate the process of data analysis, and facilitate the sharing of transcript data” (MacWhinney, 2000). The CHAT format sets out common rules for transcribers, such as the investigator, who plan to conduct their analysis and share their data using CHILDES. The major components of CHAT-originated transcripts are: 1) the file headers, 2) the main tier

and 3) the dependent tiers. The file headers open the transcript and are used only once within a file. They inform us about things such as the date and length of the recording, the names of the participants and their age, the location, and so on. The main tier (main line) is composed of words (uttered by the speakers), and special markers. Apart from the child-invented forms, most words are entered just as they are found in the dictionary. Any learner-specific forms receive a special marker attached to them on the main line. Finally, dependent tiers are lines typed below the main tiers. They contain comments, events, codes and any information of interest to the researcher. Placing them on separate tiers makes the transcript more readable, and allows researchers to include as many codes and comments as they require.

Below is an extract from a session with one of the subjects participating in this project transcribed according to CHAT.

```plaintext
(1) BEGIN
(2) LANGUAGES: en, pl
(3) PARTICIPANTS: CPL child_Polish, CEN child_English, CPR
    child_Polish+English, MPE mother_Polish+English,
    MFL mother_Polish, FPL father_Polish
(4) DATE OF SESSION: 23-Apr-08
(5) FILE: MAX43A
(6) NAME OF CHILD: Max
(7) AGE OF CHILD: 3y7.14
(8) LENGTH: 13:20
(9) HIT: nagranie wykonane przez mame dziecka: dostepne tylko w formie audio [audio format available only; recorded by the mother].
(10) MPE: a jeszcze nie opowiedziales daddy o tym, ze siæ bawiles
    ciastolinh.
(11) ENG: and you haven't told daddy that you played with playdough today.
(12) CEN: [it's -- it's --] it's a little hole and when little worms.
(13) GEN: pron[it v|be&3S det|a adj|little n|hole conj|:  coo|and
    conj|: subor|when adj|little n|worm-PL.
(14) MPL: &aha.
(15) FEN: little hole with worms coming out of it, is there?
(16) GEN: yeah.
(17) ENG: how many worms?
(18) CEN: yeah.
(19) MPE: how many worms?
(20) CEN: lots and lots and lots.
(21) GEN: [and the I c --] and then I put in a plastic box.
(22) FEN: little hole with worms coming out of it, is there?
(23) ENG: and then I put in a plastic box.
(24) CEN: yeah.
(25) GEN: you put it in a plastic box?
(26) ENG: you put it in a plastic box?
(27) CEN: yeah.
(28) ENG: how many worms?
(29) CEN: lots and lots and lots.
(30) GEN: [I --] I went to sleep and I saw just mum and I said mummy I think I want to [...]. I think what here comes babcia.
(31) MPL: mhm, tak, bo on mnie zobaczył przez okno, i ja bytem dzieci;
    mma się dalej [the mother is laughing].
(32) ENG: mhm, yes, because he saw me through the window, and I was
    picking him up on my own today, you know?
(33) GEN: [I --] I went to sleep and I saw just mum and I said mummy I think I want to [...]. I think what here comes babcia.
(34) ENG: because Granny's foot was hurting, since she scratched it and
    she had to stay at home.
```
FPL: &aha.

@MPL; no, i ja poszłam sama i mnie zobaczył przez okno i był zdziwiony, dlaczego jestem sama, nie z babcią.

%eng: so, I went on my own, and he saw me through the window, and he was surprised why I was on my own and not with gran hydraulic.

*MPL: a powiedz o tym buciku, co się stalo.

%eng: and tell [daddy] about the little shoe, what happened.

*CPL: bo był taki malutki [...].

%eng: because he was so little [...].

*FEN: you need to tell me.

*MPE; tylko daddy powiedz.

%eng: but tell daddy.

*CEN: he is very little.

*FEN: who?

*FEN: the little boy?

*FEN: [he was --] he was very little.

*CEN: but [2x] he will thinks=think [...].

*MPL: he is [...] was yours his=him.

*CEN: for him.

*MPL: pro [for] pro=him.

*MPL: tak.

*CEN: yes.

*MPL: &mhm.

*MPL: popatrz, tu powinno być [...].

%eng: look, a should be here.

*MPL: widzisz?

%eng: can you see?

*CEN: it doesn't matter.

*FPL: to zrobił=slubił jutro=lutro jak choć=co.

%eng: I'll do it tomorrow when I want to.

*MPL: przeprosi=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.

*FPL: przeprosz=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

*FPL: przeprosz=maluśka.

%eng: I'm sorry.

*MPL: nie znam, jak się mogę powyżo.

*CEN: it's fine, everyone can make mistakes.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.

*MPL: przeprosz=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.

*MPL: przeprosz=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.

*MPL: przeprosz=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.

*MPL: przeprosz=maluśka.

%eng: I'm sorry.

*CPL: nie znam, jak się mogę powyżo.

*FPL: it's fine, everyone can make mistakes.

*CEN: mama, jaki śmieś=śmiecia.

*MPL: #PRONOM=IDENT|to NOM=to #VP=biesi_FUT:SG=1:robisz=slubisz #ADV_TEMP=jutro=lutro #CONJ=TEMP jak #VI=chciać_PRZ:SG=1:choć=co.

%eng: I'll do it tomorrow when I want to.
Let us take a closer look at the headers, the tiers and the codes used in the transcript of this brief exchange between Jerzy and his parents. The top @ headers (lines 1-9 and 101) are the obligatory file headers identifying the participants, the date of the session, and other important organisational aspects of the recording. The * lines (e.g. 10, 12, 13) are the main tiers introducing the speakers' utterances. The % lines (e.g. 11, 14, 21, 22) represent a few of the 27 dependent tiers available to the transcribers. Since CHAT provides options, not requirements, there are no restrictions as to how many dependent tiers can be used in a transcript, and the number is study dependent. Some transcripts may require only the %sit tier to describe arrangements during the recording, while others will employ more tiers to give a detailed account of the context as well as enable the transcript for a pragmatic, phonological or syntactic analysis. None of the transcripts available on CHILDES to date includes codes for all those aspects in a single file. Since the aim of the present study is to investigate features of morphology and syntax, it is crucial that the utterances are coded for parts of speech and inflection. The coding has been limited to the utterances produced by the children, due to the fact that the input is not the main subject of this study and, apart from some MLU calculations, it will not be included in the in-depth analysis. The coding system for English and Polish will be described in detail later on. Project-specific speaker identification codes were introduced to distinguish between the languages used by the participants, the majority of whom are Polish-English bilinguals (see Table 16 below). A few more codes used on the main tiers of the transcript above need explanation. Square brackets (e.g., lines 13, 23, 24) are used for false starts to exclude text from the analysis - a function that is also assigned to the & sign (e.g., line 35), which is more useful for single words, such as “hm” or “aha” expressing reactions rather than content, which need to be excluded from the MLU and the frequency counts. As to the @ sign, when used on main tier beside the speaker code (e.g., line 31), it indicates that the marked speaker's utterance is not directed at the child, and is therefore automatically excluded from the input counts. Omitting the mother's explanation of the context given to the father (lines 31, 33, 36) in the transcript above would disturb the flow of the conversation, and hinder the understanding of the story told by the child. The @ code makes it possible to keep the interaction intact, at the same time excluding the utterances that do not qualify as child directed speech. The same @ sign placed on the morphological tier marks errors in children's utterances. Line 49 signals an incorrect use of the English personal pronoun, while line 80 shows an inflectional error on the masculine inanimate noun śmieć “a piece of litter”. Finally, 0 code identifies any words omitted by the child (e.g., line 89).
Table 16. Speaker codes

<table>
<thead>
<tr>
<th>CODE</th>
<th>EXPLANATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPL</td>
<td>Child Polish</td>
</tr>
<tr>
<td>CPE</td>
<td>Child Polish+English</td>
</tr>
<tr>
<td>CEN</td>
<td>Child English</td>
</tr>
<tr>
<td>IPL</td>
<td>Investigator Polish</td>
</tr>
<tr>
<td>IPE</td>
<td>Investigator Polish+English</td>
</tr>
<tr>
<td>IBN</td>
<td>Investigator English</td>
</tr>
<tr>
<td>MPL</td>
<td>Mother Polish</td>
</tr>
<tr>
<td>MPE</td>
<td>Mother Polish+English</td>
</tr>
<tr>
<td>MEN</td>
<td>Mother English</td>
</tr>
<tr>
<td>SPL</td>
<td>Sibling Polish</td>
</tr>
<tr>
<td>SPE</td>
<td>Sibling Polish+English</td>
</tr>
<tr>
<td>SBN</td>
<td>Sibling English</td>
</tr>
<tr>
<td>FEN</td>
<td>Father English</td>
</tr>
<tr>
<td>FPL</td>
<td>Father Polish</td>
</tr>
<tr>
<td>GPL</td>
<td>Grandmother Polish</td>
</tr>
<tr>
<td>BPL</td>
<td>Baby-sitter Polish</td>
</tr>
<tr>
<td>VEN</td>
<td>Visitor English</td>
</tr>
<tr>
<td>RPL</td>
<td>Relative Polish</td>
</tr>
</tbody>
</table>

3.5.2 CLAN

Once the recorded material has been transcribed in the CHAT format, it can be analysed in the CLAN program. CLAN allows researchers to perform a large number of automatic analyses, including frequency counts, word searches, co-occurrence analyses, MLU counts, interactional analyses, text changes, and morphosyntactic analysis. Particular commands need to be used to run those analyses. To illustrate the process, let us suppose that we have a file labelled `max205.cha` (the extension is added automatically to all files transcribed in the CHAT format and saved in the CLAN program). The name of the subject and his age (2 years and 5 months) are used to locate a particular file with ease. In order to obtain a frequency count in the child's utterances in this file, a following command needs to be entered in the CLAN command window:

```
freq +t*CPL max205.cha
```

An example output file:
The freq command (or any other command) can be run across a group of files. It is made possible by the so-called wildcard represented in CLAN by the * sign. A wildcard simply means that the place the * sign is put in the command can be taken by anything else. A freq command including max*.cha would run the freq count on all the files with Jerzy's in their name with the .cha extension regardless of the age marker. As to the output file, it can be shown in CLAN while the program is counting the frequency, or if a researcher wishes to save the results, it can be sent to a separate file and stored on the disc. The output file saving command is +f. Sending a frequency count from all the Jerzy’s files to one collective file can be done with the following command:
freq +t*CPL +f max*.cha

The +t and the +f options included in the freq command above are also only two out of eleven of the so-called ‘switches’ that can be used across various CLAN analysis commands, whereas the freq command is only one of a number of the so-called ‘analysis commands’ available to the researcher. Others enable:

- computing the mean length of turn (MLT)
- measuring the mean length of utterance (MLU)
- searching for complex string patterns (COMBO)
- searching used specific words in context (KWAL)
- examining parent-child repetition and expansion (CHIP)
- matching the child’s phonology to the parental model (MODREP)
- searching for previously tagged passages for further analysis (GEM)
- tracking sequences of interactional codes across speakers (CHAINS)
- computing the frequency of phonemes in various positions (PHONFREQ)
- formatting the output of FREQ for statistical analysis (STATFREQ)
- tracking the frequencies in various utterance positions (FREQPOS)
- examining patterns of separation between speech act codes (DIST)
- examining patterns of co-occurrence between words (COOCUR)
- measuring reliability across two transcriptions (RELY)
- computing the length of utterances in words (WDLEN)
- computing the Developmental Sentence Score (DSS)
- finding the longest words in a file (MAXWD)
- computing overlaps (TIMEDUR)

The list is truly impressive taking into consideration the variety of analyses it facilitates, from phonology to pragmatics, and many more. However, the list above is far from being exhaustive. There are also 28 so-called utility commands used primarily for “fixing and reformatting older files to bring them into accord with the current CHAT format or for reformatting data for use with other programs” (MacWhinney, 2000). For example, TEXTIN command converts straight text to CHAT format, while TIERORDER rearranges dependent tiers into a consistent order. On top of that, there are also six morphosyntactic commands, which is of crucial importance for the present investigation. The main command is MOR, which codes transcripts for parts of speech, enabling it for a morphological analysis and making the coding consistent, verifiable and error-free. Since the MOR function is of great importance to the present study, it will be discussed in further detail in the following subsection.
3.5.3 MOR

Undoubtedly, problems with coding schemes and cross-investigator reliability, at least for researchers working with English and a few other widely spoken languages, have been solved thanks to the development of MOR, which provides a morphological tagging list (MOR library) for the following languages: English, Italian, German, Chinese, Cantonese, Dutch, French, Hebrew, Spanish and Japanese. For those languages, CLAN offers a coded lexicon and a set of allomorphic and concatenation rules all downloadable from the CHILDES website. The lexicon comprises words classified separately according to their syntactic category. A few examples are listed below:

```
farther {[scat adj] [degree cp]} "far-CP"
old+fashioned {[scat adj][comp adj+adj]}
adamant {[scat adj]}
first {[scat adj] [num +]}
aboard {[scat adv]}
when {[scat adv:wh]}
saturday {[scat adv:tem]}
face+to+face {[scat adv][comp n+prep+n]}
do {[scat v:aux] [iry] [tense pres] [vform perf]}
do {[scat v] [iry]}
cost {[scat v] [bare yes]} "cost&ZERO"
sleep+walk {[scat v][comp n+v]}
he {[scat pro][pers 3] [num sg] [case nom]}
outa {[scat prep]} "out-prep|of"
outs {[scat npt]}
```

The examples above illustrate the parsing system according to which CLAN operates. To function, the MOR lexicon relies on three grammar files containing allomorphic and concatenation rules to specify the morphological processes of the language. The so-called ‘ar’ file (allomorphic rules) “lists the ways in which morphemes vary in shape”, while the “cr” file (concatenation rules) “lists the ways in which morphemes can combine or concatenate” (MacWhinney, 2000). The “sf” file lists special form markers, which identify forms such as neologisms, familial words, onomatopoeia, or second-language forms. All those files can be modified and altered to fit language-specific needs. Stems are stored separately, while the inflected forms appropriate for the stems are compiled at run time. Figure 2 below illustrates the structure of the MOR library:
Figure 2. MOR library
All the researcher needs to do to code new English data for morphological information is to type in the ‘mor’ command into the CLAN command window. MOR automatically generates a %mor tier for all the words in the main lines. Each word is “labelled for their syntactic category, followed by the pipe separator |, followed by the word itself, broken down into its constituent morphemes” (MacWhinney, 2000). The example dialog below illustrates CLAN coding rules:

(37) *CHI:  the people are making cakes.
    %mor:  det|the n|people v|aux|beSPRES v|make-ING n|cake-PL

Through the ‘mor’ command, the MOR program codes each word separately without looking at the context, and “provides all possible grammatical categories and morphological analyses” (MacWhinney, 2000). As a result, nearly each %mor tier may contain words that belong to two different classes:

(38) *CHI:  I want to go back.
    %mor:  pro|I v|want inf|to‘prep|to v|go adv|back‘n|back‘v|back

The caret ^ denotes the multiple possibilities for each word on the main tier. All such cases require disambiguating, usually done manually, through the POST program or Disambiguator Mode in CLAN. In comparison to the standard CLAN commands available to the researcher, the MOR feature is far more complicated and advanced. According to MacWhinney (2000),

Successful use of MOR requires a full understanding of the operation of the program, the process of lexicon building, and the use of methods for improving the morphological analysis. MOR is a complex program that is intended for the serious user who is willing to commit a large amount of time and effort in order to achieve a major improvement in analytic capabilities.

Researchers who contribute their tagging systems to CHILDES allow others to code their data automatically and reliably, which greatly enhances verifiability of their results. However, those who wish to perform analysis on the morphosyntactic level in a less frequently researched language will soon be challenged by one question: is there a tagging system or a tagging list for the language in which I work? The author of this study has been faced with this question, since a MOR library for Polish is missing from CLAN, and thus the lexicon building and the data tagging needed to be planned and processed from scratch. Entering the whole dictionary of possible word combinations into CLAN and designing codes for all the morphological features available in Polish would be an insurmountable task given the time limits that a doctoral student has at her disposal. The children participating in
this project used in total 13,487 different word types, and 101,904 words (tokens). Understandably, some words used by the three subjects were simply the same parts of speech, and coding them manually would mean copying and pasting a given code. However, 13,487 new codes would need to be manually typed into a list of codes, and then transferred to each file one by one. Coding the three corpora manually would be extremely error-prone and non-correctable; hence, it was crucial to find a way to code them automatically. The data tagging and the data analysis for this project were possible thanks to Smoczyńska's code list for Polish, which despite of being incomplete, has sped up the coding process considerably. On the basis of her own data collected throughout the years, Smoczyńska has built a tagging system for Polish sufficient for analysis in CHILDES, but providing the necessary morphosyntactic information required for investigating Polish data. Her library is not a complete dictionary of Polish words, but a collection of words used by ten children she has analysed for one of her projects. The investigator's corpora have been automatically coded by Smoczyńska, and later proofread and tidied by the investigator. Moreover, although in CLAN all the morphological information is made available by using %mor lines, this code was extended for the needs of this bilingual project, and the following codes were introduced:

- %mpl (morphology Polish)
- %mpe (morphology Polish+English)
- %men (morphology English)

Creating codes for Polish according to CLAN rules for English would be possible, but extremely time consuming. All the files, including the 'ar', 'cr' and 'sr' files and all the 'lex' files would need to be designed from scratch to include the necessary morphological information, such as gender, which are the main syntactic categories absent from English and therefore are simply excluded from the coding system. Let us compare the same Polish utterance coded according to the CLAN rules (a) for English, and Smoczyńska's code list (b) for Polish. The CLAN coding and the one designed by Smoczyńska for this utterance have been compared in Table 17:

(a) *CPL: ona chciała=ciała pobawić się.
%mor: pro|she v|want-PAST inf|to n|play.
%eng: she wanted to play.

(b) *CPL: ona chciała=ciała pobawić się.
%mpl: PRONOM: PERS: SG: 3: F|ona_NOM: ona
VI: chciał= PAST: SG: 3: P: chciał=ciała
%eng: she wanted to play.
For the personal pronoun *ona* 'she' CLAN provides the basic information (pro), whereas Smoczyńska's code also informs us about the type of the pronoun (PERS), its number (SG), person (3), gender (F) and case (NOM). Next, in CLAN the verb *chciała* 'wanted' is represented by the 'v' code plus the tense identifier (PAST) only, whereas in Smoczyńska's coding first the verb is identified as imperfective (VI), which is followed by tense (PAST), number (SG), person (3) and gender (F). Moreover, the Polish coding also caters for the child's individual pronunciation by making it possible to attach the actual child form of the verb *ciałam*. The next element is the English code *inf* which is superfluous as far as any analysis for Polish is concerned. Finally, in CLAN the verb *pobawić się* 'play' is represented by the same 'v' code as for the verb 'wanted'. The Polish coding, however, recognises not only that it is a perfective verb (VP), but also, by adding a code to the *się* element it identifies the verb as a reflexive verb. Those two pieces of information allow the researcher to differentiate this verb from the verb *bawić* 'be funny', which would be coded:

$$\#VI|bawić-INF:bawić$$

By comparison with the CLAN, Smoczyńska's coding system for Polish enables the researcher to provide a whole range of extra information vital for various types of analyses. Importantly, by include the child's forms on the main line (e.g., *ciałam*), it also equips the researcher with the power to track any cases where the child's individual phonology may cause concern by influencing the morphological aspects of the analysis. Let us consider the following example:

(c) *CPL: Magda, ja rysuję-łyśuję=takiego=takię pса=pies*.  
%mor: #NPR:F|MAGDA|SG:1|VOC:MAGDA=MAGDA  
#PRONOM:PERS:SG:1|jA NOM:jA  
#VI|rsować-PRS:SG:1:rysuje-łysiuję  
#PRAODJ:QUAL:DEM|taki|SG:M;ACC|takiego=takię  
#N:AN|pies|SG:ACC|psa=pies.  
%eng: Magda, I am drawing such a dog.
Apart from the erroneous case forms of both, the demonstrative article *takiego* ‘this’ and the noun *psa* ‘dog(M)*, the most important element to analyse in (c) is the verb *rysuj*-ę ‘draw-PRES.IMPF.ISG’ and its actual production. Some clarification is necessary before an explanation can be given. Polish verbs belong to four different conjugation classes with 1<sup>st</sup> and 2<sup>nd</sup> person singular (present tense) being the classificatory criteria. Let us compare conjugation classes I (*rysować* ‘to draw’) and III (*czytać* ‘to read’):

<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRESENT TENSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINGULAR</td>
<td>rysuję</td>
<td>rysujesz</td>
</tr>
<tr>
<td></td>
<td>czytam</td>
<td>czytasz</td>
</tr>
</tbody>
</table>

The confusion results from the fact that instead of following the rules for the conjugation class I to which the verb *rysuję* belongs, the child uses the ending *-am* (class III) and says *lysiowam*. The child’s production cannot be ignored on the basis that it is a way of omitting the nasal vowel ‘ę’, commonly pronounced as ‘e’ in Polish adult speech. By choosing to say *lysiujem*, the child might be confusing conjugation class I and IV, which is an important grammatical error, and therefore must be marked or at least listed as ‘ambiguous’. It is vital then that the child’s production is also transcribed on the morphological coding line to be taken into account when analysing the data.

### 3.6 Conclusions

A great number of methods are available for a researcher investigating child language data. For this particular topic, i.e. the acquisition of the grammatical gender by bilingual children, the longitudinal study has proven most appropriate. The observational method of sampling is not free from pitfalls, one of which is the fact that there is no strict control over how much each of the adult present would interact with the child, and consequently, to what extent one language would be used rather than the other. Understandably, more balanced output could have been expected if the recording sessions had taken place in the presence of two interlocutors, one for each language. However, the investigator had to take into account all the limitations of this project and so in deciding on the data collection method a choice was made in favour of the naturalness and spontaneity; the risk that the data might be more abundant for one language than the other was in this case inescapable. As a result, the data collection process has been supervised solely by the investigator, who arranged the vast majority of the recordings, with a number of the sessions also organised and carried out by the mothers. Overall, approximately 140 GB of data has been gathered, i.e. about 180 hours of recordings (audio and video). The recordings were labelled according to the dates of the
sessions and the age of each child. The final amount of words produced by the three main subjects and included for analysis sums up to nearly 140,000 words.

The recruiting criteria were strict and were set up to exclude overtly unsuitable candidates as well as those who were likely to choose one of the languages and abandon the other in the course of the data collection process. Unfortunately, two children out of the five recruited for the project did not produce enough data in either language to be included in the analysis. From the investigator’s observation as well as from the information given by the mothers, the children recruited for this project were healthy, talkative in both languages, and used to meeting people from different backgrounds and able to freely communicate with them.

For the analyses, the emphasis will be put on the morphosyntactic development, but we shall also look at the use and the nature of mixed utterances. There will be no reference to phonetic, phonological, semantic or pragmatic aspects. Nonetheless, since Polish has not been studied in the field of BFLA (Bilingual First Language Acquisition), it certainly contributes to the diversification within the field. The basis for investigation is the age period between 2;5-4;3 for Patrick, 2;3-4;1 for Hania and 2;5-3;10 for Jerzy. In comparison to other case studies of similar nature, e.g., De Houwer (1990), whose study is limited to the period of 8 months between the ages 2;7-3;4, this investigation is based on a much bigger age range. On the other hand, Taeschner (1983) reports on his daughter’s language development over several years, which makes this study slightly less ambitious. Nevertheless, it is hoped that the level of detail and thoroughness of the analyses will result in a reliable and useful study. In addition, the results are expected to be accurate and reliable through verification. It also needs to be stressed that the data gathered for this study have contributed to the extension of Smoczyńska’s code list, and to building a proper MOR grammar for Polish, which in the near future is likely to be publicly available via CHILDES.
Chapter Four. Data analysis

4.0 Introduction

The previous chapters either provided theoretical background for this study, or dealt with descriptions of methodological procedures undertaken to carry out the data collection and coding process. The aim of this chapter is twofold. First, by assessing and comparing the children’s very early data, it illustrates the initial stages of the children’s linguistic development and introduces their linguistic profiles. Second, it analyses the data. The analyses are based on mixed utterances and monolingual Polish utterances. Monolingual English utterances have been excluded from the analysis due to the insufficient amount of data pertaining to grammatical gender.

4.1 Language profiles

It is a standard procedure for a child language researcher to begin their investigation by establishing the level of grammatical complexity of their subjects’ language. In any study involving more than one subject, it is vital to determine to what degree they share their skills in language perception and production. Human language development is a rather individual phenomenon; therefore, biological age on its own cannot function as a basis for comparing young speakers’ linguistic development. Linguists and clinicians have developed various language growth measuring tools, the majority of which have been created with English in mind. The underlying principles of those tools followed the trends in linguistics that were current around the time the tools were created. With time, linguists have developed new theories. Subsequently, the language growth measuring indices have had to be modified and updated to respond to those new, improved approaches to language. Based on speech samples taken from the CHILDES database, Pan (1994) analysed longitudinal transcripts from 48 children at the age of 14, 20 and 30 months, and generated language profiles for five children chosen from the same corpus. After it had been edited to reflect changes in
programmes and the database, Pan’s work was later included in the updated CLAN manual (2008) for the CHILDES scheme. The five language profiles generated by Pan were based on the following basic measures: MLU50 (mean length of 50 utterances), MLU5 (mean length of five longest utterances), TTR (type-token ratio), and MLT (mean length of turn). To compare her results with the ones of the peers from the corpus after the above measures had been calculated, the author generated z-scores for each of the five children by using the means and the standard deviation for each measure for the whole corpus at 20 months.

A similar analysis generating language profiles for Hania, Jerzy and Patrick has been conducted for this study. However, despite the fact that the inspiration comes from Pan’s work, some alterations have been introduced to the procedure: firstly, MLU has been calculated in words, not in morphemes. Secondly, rather than on 50 utterances, MLU has been calculated on the child’s utterances recorded during a given month. Finally, there is no comparison between the subjects of this study and their peer group, as there is no Polish-English corpus with speech samples from 48 children available for such an analysis. Therefore, any comparisons are made between the three participants only. Additionally, a group of three is too small to offer definite answers as to the syntactic, lexical and conversational abilities of normally developing bilingual children, therefore any conclusions are to be understood as tendencies, not generalizations. The three case studies described here serve as a window through which we can look only to see three various views on similar aspects.

MLU
Based on the MLU values obtained from the three children, Adam, Eva and Sarah, Brown (1973) constructed a 5-stage ladder of linguistic development: stage I (1.0-2.0), stage II (2.0-2.5), stage III (2.5-3.0), stage IV (3.0-3.5) and stage V (3.5-4.0). Interestingly, Brown was aware of both advantages and also disadvantages of MLU (1973: 53-4):

The mean length of utterance (MLU) is an excellent simple index of grammatical development, because every new kind of knowledge increases length: the number of semantic roles expressed in a sentence, the addition of obligatory morphemes, coding modulations of meaning, the addition of negative forms and auxiliaries used in interrogative and negative modalities, and, of course, embedding and coordinating [...] By the time the child reaches stage V, however, he is able to make constructions of such great variety that what he happens to say and the MLU of a sample begin to depend more on the character of the interaction that on what the child knows, and so the index loses its value as an indicator of grammatical knowledge.

40 For more information about the three children see § 3.0 of this thesis.
The reason why the usefulness of MLU beyond 4.0 decreases is the fact that rather than adding new structures, the child begins to reorganise the utterances by building clauses, and using more sophisticated devices, such as ellipsis, which often results in shorter utterances. MLU at that stage would provide misleading results suggesting that the child’s linguistic development is deteriorating rather than improving. In the following sections, MLUw (in words) plus two other measures will be generated with the use of CLAN to generate Language Profiles of Patrick, Jerzy and Hania. The goals are to use the data from these normally developing children to inform us as to their performance, and to provide an intragroup comparison between them.

Obtaining results through CLAN for Polish data required overcoming certain obstacles. The first was the need for a clear-cut set of morphemicization rules for Polish included in the Polish MOR library. As has been explained in Chapter Three, there is no MOR library for Polish, and building one to use in this project would be unrealistic, taking into consideration its time limits. Consequently, following Parker and Brorson’s (Parker and Brorson, 2005) conclusion regarding the reliability of MLU and MLUw, a decision was made to provide the MLU measurements for Polish on the word level, rather than the morpheme level:

MLUw can be used as effectively as MLUm in the measurement of gross language development. MLUw is a more reliable measure of gross language development than MLUm due to the fact that arbitrary decisions regarding morpheme assignment are eliminated. Furthermore, MLUw is a more effective measurement as it can be used more readily and reliably across various languages.

Those arbitrary decisions regarding morpheme assignment were avoided in both cases, English and Polish. Both languages were measured in MLUw, despite the fact that there are rules for morpheme counts for English. One of Brown’s instructions for calculating MLU is to choose 50-100 intelligible utterances, ideally starting with the second page of the transcript. This is believed to increase the reliability of the measurement, as the child usually takes a few minutes to “warm up”. This instruction is relevant for manual counts, but not for the automatic counts provided by the CLAN program. Moreover, counting words on a sample of 100 utterances can be extremely unrepresentative due to very varied dynamics of each session with the child. At one point, a conversation about animals or new toys will be interesting to a child, but a month later the same toys will no longer be as popular, which may result in the child’s responses being shorter and limited to one-word or two-word utterances. Additionally, the transcripts from all the sessions recorded within a given month were put together as one transcript and labelled collectively (e.g., Hania 3;1), rather than as individual session recorded on particular days during that month (e.g., Hania 3;1.2, Hania
Such an approach makes tracking the monthly linguistic growth much easier. It also means that choosing a sample of 100 utterances from the beginning of the whole transcript renders results that might be significantly lower than if the calculations were made based on the last 100 utterances in the monthly transcript. Miller and Chapman (1981: 158) cautioned the reader that MLU is "sensitive to contextual variables such as the nature of the interaction", which also means that it is extremely difficult to randomly pick 50 most representative utterances from the transcript. Therefore, in this project, it was decided against measuring the MLUw on a selected number of utterances, but rather, it was calculated on the entire material collected during a given month of the child's linguistic development. Finally, the bilingual character of children's utterances required a unified measuring approach; hence, it was necessary to run the MLU in words on the Polish, the English and the mixed utterances, despite availability of the English morphemicization rules.

The language profiles are generated for the age of 2;5 (29 months), which is when the children began their recording sessions. The first CLAN analysis we need to perform for each child is MLUw. By default, the MLU program excludes the strings xxx, yyy denoting unintelligible utterances, and www, which are fragments that were purposefully left untranscribed (e.g., a conversation between adults). Utterances that contain any of those strings are also excluded, even if other parts of the utterance are fully transcribed. Moreover, excluded are also any strings immediately preceded by one of the following symbols: 0, &, +, -, #, ;, or $ (see §3.5.1). The following command has been used to compute the MLU for the children:

\[ \text{mlu } +t\ast \text{C* } @ \]

The name of the program has to come first in the command, and it can be followed by any command/switch desirable. The +t switch means "analyse only". The first (*) is just a part of the speaker code used according to the CHAT transcribing conventions, whereas the second (*) symbolizes a wildcard. In this case, it is used to include all the child's utterances: in Polish (CPL), in English (CEN) and mixed utterances (CPE). Finally, the @ symbol appears in the command window when the input file(s) have been selected from a list of files rather than specified manually. The MLUw output for the three children aged 29 months (computed on the basis of a 40-minute recording per child) is as follows:

<table>
<thead>
<tr>
<th>Table 18. MLU at the age of 29 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Hania</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>MLU</strong></td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
</tr>
<tr>
<td><strong>Number of utterances</strong></td>
</tr>
<tr>
<td><strong>Number of words</strong></td>
</tr>
<tr>
<td><strong>Number of repetitions</strong></td>
</tr>
</tbody>
</table>
As we can see, at the beginning of the recording process, the children were all at the Stage I of their linguistic development. However, their MLUs differed hugely. At the age of 29 months, Hania's MLU was only 1.0. She produced single word utterances in almost all cases, 33% of which were repetitions and 67% was Hania's own production. In comparison, Patrick's MLU was 1.7, and repetitions were almost non-existent. Interestingly, Patrick produced nearly the same number of utterances as Hania, but about 40% more words, which indicates that his utterances were definitely two-word and longer. Jerzy's MLU was the highest of all at the age of 29 months, and his utterances were the longest as well. The percentage of the repetitions in Jerzy's and Patrick's speech was 3.6% and 0.8% respectively, which is minimal in comparison to Hania's 33%. The standard deviation gives us some indication of how variable the child's utterance length is. At this point, Patrick and Jerzy were comparable in their linguistic stage of development. They could build multiword utterances, which are nearly always their own independent production.

MLU5

The next CLAN analysis to be performed is the MLU5, i.e. the mean length in words on each child's five longest utterances. The following command string has been used:

```
maxwd +t*C* +g2 +c5 +d1 @ | mlu +t*C* > ame205.ml5
```

There are two commands in this string. First, the maxwd locates and measures either the longest word or the longest utterance in a file. The +g2 switch identifies the longest utterance in terms of words, the +c5 switch tells the program how many utterances to find, and the +d1 string tells maxwd to send output to the output file in CHAT format to make it available for analysis by other CLAN programs. The second command in this string begins with the pipe symbol |, which sends the output of the first command to the next command, i.e. mlu. Finally, the redirect symbol > followed by the output filename and extension specifies where the final output file is to be saved.

<table>
<thead>
<tr>
<th>MLU5</th>
<th>Hania</th>
<th>Patrick</th>
<th>Jerzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>1.8</td>
<td>4.8</td>
<td>6.0</td>
</tr>
<tr>
<td>Number of utterances</td>
<td>0.400</td>
<td>0.748</td>
<td>1.549</td>
</tr>
<tr>
<td>Number of words</td>
<td>5</td>
<td>24</td>
<td>30</td>
</tr>
</tbody>
</table>

There is a considerable gap at this initial stage between Hania, who produced only nine words in the total of five longest utterances, and the two boys, who produced between 4.8-6.0 words in each of their five longest utterances. However, Hania's initial slow language
learning tempo was only short term, as she later produced utterances that are equally complex in terms of grammar, but also were the most correct in terms of gender, number and other linguistic features.

TTR
The third CLAN analysis is to compute Type-Token Ratio, for which the FREQ program is required. Just like MLU, FREQ (frequency) ignores strings such as xxx, www, or any of the symbols mentioned earlier. Here it is not important how many words the child used in a certain number of utterances, or whether the child produced any plural forms. TTR shows how many different vocabulary items were used (word-types), and how many forms of the same word-form were used (tokens). For example, forms such as umiesz ‘can.2SG.PRES’ and umiem ‘can.1SG.PRES’ are supposed to be counted as two tokens of the word-type umieć ‘be able to’. Unfortunately, without the MOR library for Polish, CLAN treats each unique item as a separate type. Consequently, had the count been done automatically, umiem and umiesz would have been counted as two different types, rather than two tokens of the same type. The result yielded would be utterly unrepresentative, thus the TTR has been counted manually. A frequency list for each child has been obtained with the following command:

freq +t*C* @

The list has been carefully analysed and tokens of the same word-types marked as such. Table 20 presents the TTR results for the three participants:

<table>
<thead>
<tr>
<th></th>
<th>Hania</th>
<th>Patrick</th>
<th>Jerzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTR</td>
<td>0.31</td>
<td>0.23</td>
<td>0.29</td>
</tr>
</tbody>
</table>

MLT
The final analysis performed to complete the children’s linguistic profiles was MLT (mean length of turn) for both the children and their adult interlocutors. Despite the absence of semantically and/or grammatically valuable material, and unlike MLU, this program includes the strings xxx, www, as they constitute turns in the conversation. For the same reason, strings normally excluded by MLT have also been added to the count: repetition marker [/], and all the “empty” responses: &yy, &ee, &aa, &mhm, &hm, &oo. The MLT has been first computed for the child and then for the mother. The following TTR command string has been used:

mlt +t*C* +s+"&" +s+"</>" @
The command is different for adult interlocutors since no repetitions are involved. Pan’s (1994) MLT count involved the child and the mother only. In this project, the investigator, the fathers, carers and siblings would often be involved, and they should naturally be included in the turn-taking measurements. Since it does not matter which person other than the child is taking the turn, the results for all interlocutors taking part in the conversation with the child were summed up. The number of words over turns was chosen rather than the number of utterances over turns, following Pan’s reasoning according to which “words per turn is likely to be sensitive for a somewhat longer developmental period” (1994: 42). All words produced by the child’s interlocutors were added and divided by the sum of their turns. Next, the child’s MLT (words over turns) was divided by the joined MLT of the other speakers. The results are presented in Table 21 below. Similarly to the earlier measurements, in comparison to the boys, at this point Hania came out as a weak conversation participant. Her ratio of words over turns was 48% smaller than Patrick’s and 52% smaller than Jerzy’s, who shared the biggest part of the conversational load in this group.

<table>
<thead>
<tr>
<th></th>
<th>Hania</th>
<th>Patrick</th>
<th>Jerzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s MLT</td>
<td>1.095</td>
<td>2.072</td>
<td>2.320</td>
</tr>
<tr>
<td>Adult’s MLT</td>
<td>4.5176</td>
<td>4.8875</td>
<td>4.5558</td>
</tr>
<tr>
<td>MLT ratio</td>
<td>0.2423</td>
<td>0.4245</td>
<td>0.5091</td>
</tr>
</tbody>
</table>

4.2 Data

4.2.1 Polish data

A likely route to learning a gender system is believed to begin with recognising the patterns operating in a given language rather than memorizing every single noun with its gender (Corbett, 1991: 82). Monolingual Polish children begin to recognise those patterns even at the age of 18 months, when they produce the first morphologically “interesting” phrases, involving features such as gender, case and number. The three children studied here are older – the first recordings date to the age of 2;4 (Jerzy), 2;5 (Hania), and 2;5 (Patrick). Such a delay does not necessarily mean that the important initial data was missed. The Language Profiles presented in §4.0.4 show that on entering the project all the children were at Stage I of their linguistic development (MLU between 1.0-2.0). It means that the discrepancy between the age when monolingual Polish children are reported to begin their morphological adventure and the children studied here is not significant.
Analysis 1 Diachronic error shift

According to Rowland, children do not learn errors imitatively from adults, and for that reason errors are believed to “provide insight into the child’s grammatical system at a given point in development” (2006: 862). Researchers, who have investigated the validity of estimates of error rates based on small samples report that “overall error rates underestimate the incidence of error in some rarely produced parts of the system” and that “analyses on small samples were likely to substantially over- or underestimate error rates in infrequently produced constructions” (Rowland and Fletcher, 2006). Importantly, the same researchers suggest caution when basing arguments about the scope and nature of errors in children’s productions on analyses of spontaneous speech:

Speech, even adult speech, tends to be made up of a small number of words that occur often (e.g. the wh-words what and where, the verbs do and be) and a much larger number of words that occur far less often (e.g. why, when, bounce, gobble). The high frequency items are more likely to be represented in any given sample than the low frequency ones. Thus, the traditional measure of lexical specificity – demonstrating that a significant proportion of a child’s utterances can be accounted for by a small number of lexically specific frames – is confounded by the fact that a small number of highly frequent utterance types are likely to account for a large amount of the data anyway. Analyses based on these samples may, then, underestimate the variety and productivity of children’s speech (Naigles, 2002). (Rowland and Fletcher, 2006: 861)

Those concerns are certainly important if we are investigating infrequent features of language, the occurrence of which is difficult to predict or evoke. Fortunately, grammatical gender in Polish is marked on many different elements and, if necessary, it is not difficult to elicit data on the child’s gender-marking skills. In Polish NPs, grammatical gender is obligatory, i.e. children must make a specific gender choice when building an NP, or a VP in the past tense. Therefore, concerns about capturing only a small number of highly frequent utterances in spontaneous recordings are not applicable in this study. However, there is a potential danger of missing temporary gender errors, i.e. those, which last only a short period of time in the children’s speech. It is impossible to point to specific errors due to the individual nature of the process: one feature of gender marking (e.g., animacy) may be clear early on to one child, but may not be discovered till later by another.

Researchers have often suggested that “many of the types of error we might plausibly expect to see are extremely rare or even nonexistent in children’s speech” (Rowland and Fletcher, 2006: 863), which has led to the conclusion that children acquire adult-like competence in language production early and quickly, and that very young children have a sophisticated knowledge of a range of syntactic constructions. Polish researchers have observed that Polish-speaking children typically learn grammatical gender
early and nearly effortlessly, which suggests either that there are only few gender errors, or that due to the nature of research those infrequent errors were missed. The author possesses knowledge of the linguistic development of a monolingual Polish-speaking girl, Pola, who at the age of 2 was about 80% productive in her case system with the ability to form agreement with nouns of all genders and numbers. Understandably, children in general do make errors in their language production, but the question is: can we observe any patterns in their acquisition of a particular linguistics area? Spreng (2004) discusses the error patterns in the acquisition of German plural morphology and shows that gender functions as a cue to learning the right plural marker in German. Based on her observation, she concludes that “errors demonstrate in a predictable way what children use as cues for acquisition” (Spreng, 2004: 147). The main locus of gender errors examined here is gender agreement, but all charts also include personal pronouns, which reflect the mapping process of natural and grammatical gender. In the following paragraphs, the diachronic spread of gender errors found in the database will be investigated with the hope of uncovering interesting details of the gender learning process by Polish-English children. The data allows for measurement of relative error type proportions, i.e. it does not compare error rates to the overall rates of correct agreement forms. Unavoidably, language production is context-bound; hence, errors found in the database may or may not reflect the children’s lack of certain linguistic knowledge. They may be there because the conversations that were held that day focused on a particular game or event, and more forms of particular gender were needed. The discourse is an organic part of data collecting process; it is a requirement, but at the same time, it determines the number of forms, and also errors that occur, which should also be taken into account when we think about the nature of child language research.

My error database consists of a total of 25 error types, which fall into 4 classes: Classes 1 and 2 include errors in the singular, Class 3 contains errors in the plural, and Class 4 includes errors across numbers, i.e. errors both in number and gender. Errors involving animacy are listed in Class 2 separately from other singular errors.

<table>
<thead>
<tr>
<th>Table 22. General error classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Errors within singular</strong></td>
</tr>
<tr>
<td>Class 1</td>
</tr>
<tr>
<td>a f/n/m &gt; f/n/m</td>
</tr>
<tr>
<td>Class 2</td>
</tr>
<tr>
<td>a f/n/m &gt; f/n/m</td>
</tr>
<tr>
<td>b m (in)an &gt; m (in)an</td>
</tr>
<tr>
<td>c m (in)an &gt; f/n</td>
</tr>
<tr>
<td><strong>Errors within plural</strong></td>
</tr>
<tr>
<td>Class 3</td>
</tr>
<tr>
<td>(non)mp &gt; (non)mp</td>
</tr>
<tr>
<td><strong>Errors across number</strong></td>
</tr>
<tr>
<td>Class 4</td>
</tr>
<tr>
<td>a (non)mp &gt; f/n/m</td>
</tr>
<tr>
<td>b f/n/m &gt; (non)mp</td>
</tr>
<tr>
<td>c m (in)an &gt; (non)mp</td>
</tr>
</tbody>
</table>

> indicates the direction of replacement (left to the symbol are expected correct genders; right to the symbol are genders actually produced)

1 = f, n, m is exchanged for any of the other two
2a = using m inanimate instead of m animate and vice versa
2b = m (in)animate is used instead of f or n
2c = f, n is used instead of m (in)animate etc.
Although based on the more fine-grained classification of the error types (below) we can only measure relative error type proportions (e.g., is the error $f>m$ made more or less frequent than $m>f$ when the children are young or older, etc.), the children’s individual patterns are observable and will serve as the basis for drawing conclusions. The distinction between error type $f/n > m$ ($f/n$ is expected, but $m$ is produced) and $f/n > m$ (in)animate is detectable in the accusative case due to the presence of the animacy feature. Hence, all accusative examples involving masculine agreement forms have been classified as involving masculine animate (if the agreement target is showing a masculine animate ending), or masculine inanimate (if the agreement target is showing a masculine inanimate ending). Agreement examples in accusative where either feminine or neuter was involved are not included in Class 2 for obvious reasons (the lack of the animacy feature). A fine-grained error type classification is presented in Table 23:

<table>
<thead>
<tr>
<th>CLASS CODE</th>
<th>ERROR CLASS</th>
<th>ERROR TYPE</th>
<th>ERROR NR</th>
<th>ERROR TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$f/n/m &gt; f/n/m$</td>
<td>$m &gt; f$</td>
<td>294</td>
<td><strong>708</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$f &gt; m$</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n &gt; m$</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$m &gt; n$</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$f &gt; n$</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n &gt; f$</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>(in)animate</td>
<td>$m &gt; n$</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>(in)animate</td>
<td>$f &gt; n$</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>(in)animate</td>
<td>$n &gt; f$</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$f/n &gt; m$</td>
<td>$f &gt; m$</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$f &gt; m$</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n &gt; f$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>(in)animate</td>
<td>$f &gt; n$</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>(in)animate</td>
<td>$n &gt; f$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(non)animate</td>
<td>nonanimate</td>
<td>116</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nonanimate</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>(non)animate</td>
<td>nonanimate</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>(non)animate</td>
<td>nonanimate</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$f/n/m &gt; (non)mp$</td>
<td>$m &gt; f$</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n &gt; f$</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Rather unsurprisingly for the age group considered here, the majority of errors are within the singular. The large number of errors involving animacy (Class 2), albeit unexpected, is very welcome, as it suggests that the children are not simplifying the masculine gender by avoiding the feature of animacy. Class 4 shows that singular genders are called upon when the children have problems with number. Errors across number constitute only c. 7% of all
gender errors, which without a more thorough analysis of number errors tells us only that between the age 2;5-4;3, bilingual Polish-English children make very few errors resulting from gender choices pertaining to the plural. Another observation worth mentioning is that children make twice as many errors within the plural itself than across number, which may mean that at some point they stop being limited by the knowledge of singular genders only and concentrate on sorting out the rules for the masculine personal/non-masculine personal agreement. The frequency of error types is listed in Table 24. As it is hardly reliable to regard one-, three- or seven-member groups as representative of a certain error type, particularly after considering the length of the data collection process (two years) and the resulting amount of data, groups containing errors of frequency lower than 1% (fewer than ten examples) have been excluded from the figures. The most frequent replacement gender in the monolingual Polish data is masculine. Nonetheless, it would be hasty to suggest that masculine takes over other genders all or most of the time.

<table>
<thead>
<tr>
<th>ERROR TYPE</th>
<th>ERROR NR</th>
<th>% of total errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>m &gt; f</td>
<td>294</td>
<td>27.0%</td>
</tr>
<tr>
<td>f &gt; m</td>
<td>263</td>
<td>24.0%</td>
</tr>
<tr>
<td>mp &gt; nonmp</td>
<td>116</td>
<td>11.0%</td>
</tr>
<tr>
<td>f &gt; m inan</td>
<td>57</td>
<td>5.5%</td>
</tr>
<tr>
<td>n &gt; m</td>
<td>54</td>
<td>5.0%</td>
</tr>
<tr>
<td>nonmp &gt; m</td>
<td>51</td>
<td>4.5%</td>
</tr>
<tr>
<td>m inan &gt; m inan</td>
<td>48</td>
<td>4.4%</td>
</tr>
<tr>
<td>m &gt; n</td>
<td>46</td>
<td>4.4%</td>
</tr>
<tr>
<td>f &gt; n</td>
<td>38</td>
<td>3.5%</td>
</tr>
<tr>
<td>nonmp &gt; mp</td>
<td>36</td>
<td>2.7%</td>
</tr>
<tr>
<td>n &gt; m inan</td>
<td>17</td>
<td>1.5%</td>
</tr>
<tr>
<td>nonmp &gt; n</td>
<td>16</td>
<td>1.5%</td>
</tr>
<tr>
<td>n &gt; f</td>
<td>13</td>
<td>1.3%</td>
</tr>
<tr>
<td>m inan &gt; m inan</td>
<td>12</td>
<td>1.1%</td>
</tr>
<tr>
<td>nonmp &gt; f</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>m inan &gt; f</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>mp &gt; m</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>f &gt; m inan</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>f &gt; nonmp</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>m inan &gt; n</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>mp &gt; f</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>n &gt; mp</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>n &gt; m inan</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

In the course of the two-year recording period, the three participating children produced gender errors with varying frequencies. As is in the very nature of naturalistic research, the length of each child's monthly recordings vary, therefore some periods are directly comparable across all three participants, while others are comparable between only two of them. The frequency of gender errors, however, is an individual trait of each child.

Replacement gender is used instead of the adult gender.

124
Figure 3 shows three trends, one of which has a very low variance between the first and the second half of the recording, and is therefore statistically insignificant ($R^2 = 0.04$ for Patrick). The remaining two trends are significant ($R^2 = 0.37$ for Hania, $R^2 = 0.65$ for Jerzy), and indicate the existence of an underlying pattern in the error production (the level of error production is higher in the second half of the recording than in the first half).

Figure 3. Children's error frequencies
(numbers 1-19 mark months of participation in recordings)
Let us take a closer look at the change over time in each error type found in the raw database, in which we aim to see the children's progress in the learning of the gender system. The charts to follow present shifts in errors and represent gender-learning tendencies, as well as the children's preferences when choosing erroneous gender in agreement situations. Some individual observations will also be offered. The figures illustrating the shifts in the gender preferences are generated from two perspectives: 1) showing genders replaced by other genders; 2) showing genders as the replacing (preferred) genders. The overall aim is to observe changes over time in the general and the individual gender preferences, and link them to the gender choices made in the mixed NPs. It is expected that there is a correlation between the children's gender preferences in the mixed NP and their gender errors in the Polish database. More specifically, the particular genders used more actively in mixed NPs are expected to correlate with them also being chosen more often as replacement genders in the Polish database. If this prediction is confirmed in the database, it will suggest that mixed NPs (and ultimately, English influence) either mirror the gender choices from the Polish agreement situations, or vice versa: choices made in the mixed NPs have bearing on the children's errors made in the Polish data.

**Errors within the singular**

**CLASS 1**

\[f/n/m > f/n/m\]

This group constitutes 65% of all the errors, and is therefore the largest error group in the database. Starting with the feminine gender, the immediate observation based on Figures 4-6 is that masculine replaces feminine far more often than the neuter does. A closer look at Figures 4-6 also shows that around the age of 2;8 and 2;10, the number of neuter replacing feminine is growing while the number of masculine replacing feminine is decreasing. Around the age of 3;5-3;6 there is an opposite tendency, namely masculine is used as a replacement for feminine more often, whereas the frequency of using neuter as a replacement gender keeps dropping until the age 3;9. The numbers for neuter are particularly difficult to interpret, because they are very low. It is noticeable, however, that between 2;11-3;9 neuter is gradually falling, while masculine keeps rising. The gradual drop in neuter replacing feminine suggests that when contrasted with feminine, neuter is becoming a less frequent choice, because the agreement rules for neuter, not feminine, are becoming clearer.
to the children\textsuperscript{42}. Another intriguing observation concerns the low number of both masculine and neuter replacements for feminine around the age 3;4. The amount of data recorded for those months is ample, so the risk of drawing conclusions based on insufficient data can be safely excluded. The falling tendency for feminine gender to be replaced by any other gender suggests that the children found feminine easy to deal with in agreement situations around that particular time. In contrast, the low number of errors at the age 3;9 does not result from the fact that the children are particularly skilled with feminine agreement at that age. Rather, one or two children could not participate in recordings and their data were limited that month. Finally, the reader's attention should certainly be drawn by the rocketing number of errors in feminine around the age 3;10. The length of the recorded material for the age 3;10 equals that of 3;7, and is nearly half that of 3;2, therefore, we can safely exclude the possibility of having too much data for the age 3;10. For some reason, one, two, or all the children used masculine quite often where feminine was expected. It overlaps with the high number of errors in masculine where feminine was used instead, as shown in Figures 4-6. It is posited here that, according to the data, around the age 3;10 the gender agreement for these two genders is still not acquired.

Figure 4. Feminine replaced by masculine or neuter (Hania)

\textsuperscript{42} This observation will be further contrasted with Figure 9, in which neuter replaces masculine, and a slightly different pattern is observed.
Figure 5. Feminine replaced by masculine or neuter (Patrick)

Figure 6. Feminine replaced by masculine or neuter (Jerzy)

Examples:

(39)  
Byl-a—był-o\textsuperscript{34}  
cytrynka.  
be.3SG.PST-F=3SG.PST-N  
lemon(F)  
'There was a lemon.' [Jerzy 2;11]

(40)  
Ona=on= byl-a—był*  tak-a—tak-i*  krew.  
she=he be.3SG.PST,F=3SG.PST,M such-F=M blood(F)  
'lt was such blood.' [Pat 3;9]

(41)  
To  byl-a—był-o*  fajn-a=fajn-e*  książka.  
it be.3SG.PST-F=3SG.PST,N cool-F=N book(F)  
'It was a cool book.' [Jerzy 3;10]

(42)  
Tylko co  tamt-a=tamt-o*  rzecz  byl-a—był-o.*  
only what that-F=N thing(F) be.3SG.PST-F=3SG.PST,N  
'Only what was that last thing.' [Pat 4;1]

\textsuperscript{34} The correct gender choice is on the left of the `=' sign, and the actual gender used is shown on the right.
The examples provided above show the agreement situations in which feminine gender is replaced with masculine or neuter. The agreement targets vary, as do the nouns involved. The diversity of nouns found in the database (here only exemplary) shows that the children’s gender errors in feminine involve diminutive nouns (cytrynka ‘lemon(F)’), regular nouns (książka ‘book(F)’), and nouns ending in a consonant (krew ‘blood(F)’, rzecz ‘thing(F)’). Examples 40 and 42, which come from the same child, demonstrate how much time children may need to learn the following rule: a small group of nouns ending in a consonant declines according to declension III and takes feminine agreement. A quick look at utterances containing correct agreement during the 13-month gap between ex. 40 and 42 reveals that feminine nouns ending in a consonant take both feminine and masculine agreement (even within the same conversation). We saw the relevant noun declensions and associated assignment in §1.2.2.2 (Chapter One).

We now turn to masculine agreement, in which instead of masculine, the children use feminine or neuter. Around 2;10-2;11 and 3;5, neuter is used as a replacement gender more often than at other ages. Interestingly, around that time, neuter is also used a few times in mixed NPs. The general data is insufficient, however, to suggest that at these points in time the children focus on the neuter gender more than on the other genders. Feminine starts replacing masculine around the age of 2;6, which is at least two months after the children use masculine to replace feminine. The process of clarifying the rules of feminine gender assignment is likely to take place between 3;0 and 3;7, which is when feminine is used most actively as a replacement gender for masculine.

As a replacement gender, neuter is used very rarely. Nonetheless, the fact that more instances are recorded when it is contrasted with masculine than with feminine (Figure 8) indicates that the assignment rules of neuter are easier to distinguish from the assignment rules of feminine than from those of masculine. Its role as a replacement gender for feminine diminishes gradually from the age 2;11, but for masculine, it is used actively, particularly at 2;10-2;11, 3;4-3;5 and 3;10. This can be explained based on the similarities found among the inflectional endings that constitute declensional classes for masculine and neuter, and which are found on both controllers and targets. In most cases, except for nominative and accusative, inflectional endings given to masculine and neuter controllers are identical, as are the inflectional endings given to the targets. This may be confusing to children, who are trying to learn the gender of a noun by making a correct judgement as to the agreement rules between the controlling noun and the target. A child learning the rules of gender assignment is likely to make errors due to the declensional similarities, which may lead to errors in gender agreement. Because of the closeness of the declensional paradigms, the confusion between masculine and neuter is expected to last longer than the confusion between feminine
and neuter. Below is an extract from Table 4 (see Chapter One), showing the similarities between the inflection endings discussed:

<table>
<thead>
<tr>
<th></th>
<th>MASCULINE</th>
<th>NEUTER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>ten student</em>  *student*</td>
<td><em>ten kot</em> <em>cat</em></td>
</tr>
<tr>
<td>GEN</td>
<td><em>tego studenta</em></td>
<td><em>tego kota</em></td>
</tr>
<tr>
<td></td>
<td><em>temu studentowi</em></td>
<td><em>temu kotu</em></td>
</tr>
<tr>
<td>DAT</td>
<td><em>temu panu</em></td>
<td><em>temu winu</em></td>
</tr>
<tr>
<td>INS</td>
<td><em>tym studentem</em></td>
<td><em>tym kotem</em></td>
</tr>
<tr>
<td>LOC</td>
<td><em>tym studenci</em></td>
<td><em>tym kocie</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>tym winem</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>tym winie</em></td>
</tr>
</tbody>
</table>

Figure 7. Masculine replaced by feminine or neuter (Hania)

Figure 8. Masculine replaced by feminine or neuter (Patrick)
Examples:

(43) To już byl=był-o* banan. [Jerzy 2;11]
   it already be.3SG.PST.M=3SG.PST-N banan(M)
   'Banana has already been (used).'

(44) Tak-i=tak-a* aparat star-y. [Han 3;1]
   that-M=F camera(M) old-M
   'That old camera.'

(45) Mama, gdzie jest moj-o=moj-a* maszynista? [Pat 3;0]
   mum where be.3SG.PRES my-M=F train.driver(M)
   'Mum, where is my train driver?'

(46) Ten byl=był-o* zepsut-y=zepsuta* już [latawiec]. [Jerzy 3;10]
   this be.3SG.PST.M=3SG.PST-N broken-M=F already [kite]
   'It [kite] has already been broken.'

(47) Zatrzymał-em=zacymal-am* żeby powiedzieć [Jerzy 3;10]
   stop.1SG.PFV.PST-M=1SG.PFV.PST-F to tell
   ci kocham.
   you.DAT love.1SG.PRS
   'I stopped to tell you I love (you).'

An NP involving a gender error on the demonstrative pronoun, while no error is made on the adjective (ex. 44) raises the following question: is the distribution of gender errors in the monolingual Polish data similar to or different from the one in the mixed utterances? In ex. 44, the two forms (correct and incorrect) appear in the attributive position, contrary to ex. 46, in which two gender errors are produced, one in the attributive and the other in the predicate position. In ex. 43, Jerzy confuses neuter and masculine on the past tense 3rd person verb, and repeats the same mistake nearly a year later. In ex. 46, he confuses
masculine with feminine in the same utterance, thus producing a "double gender error". The involvement of different parts of speech in gender errors will be further discussed in §4.2.2.

We finally turn to neuter as the replaced gender. Both feminine and masculine are used as replacement genders (masculine is used more often than feminine). Although low numbers prevent us from providing useful observations as to tendencies in the error rates in neuter, two stages are particularly interesting: around the age 3;1 masculine and feminine replace neuter equally often, but between 3;2 and 3;7, only masculine is used as a replacement gender. This situation maps the earlier observations for feminine being replaced by neuter, which gradually becomes less often used as a replacement gender. Here, feminine also stops being the replacing gender, likely because the rules of the feminine and neuter agreement are becoming clearer to the children. The high frequency of feminine replacements for masculine between 3;0-3;7 mentioned above, contrasted with low frequency of feminine replacements for neuter between 3;2-3;7, provide more evidence to support the claim that feminine and neuter are beginning to be regarded as separate genders especially between 3;0-3;7. Here are a few examples of neuter replaced by masculine and feminine:

Examples:

(48) *polietal-o=polietal-a*.
fly.3SG.PFV.PST-N=3SG.PFV.PST-F
*jajko.* egg(N)
'The egg has flown (away).'

(49) *t-o=t-en* mal-o=mal-y*
drezwo [...].
this-N=M small-N=M tree(N)
'This small tree (s/he) ... '

(50) To jest tak-ie=tak-i*
masho.
this be.3SG.PRES such-N=M butter(N)
'This is such butter.'

Feminine replaces neuter at an early stage (ex. 48), and masculine replaces neuter at later stages (ex. 49 and 50). The three figures below provide a comparative image of how genders within singular are activated as replacement genders throughout the recording period between ages 2;4 to 4;3 (here we also include neuter, despite the low numbers). Some individual observations regarding errors in Class 1 are discussed below.
Figure 10. Feminine as the replacing gender

<table>
<thead>
<tr>
<th></th>
<th>2;5</th>
<th>2;6</th>
<th>2;7</th>
<th>2;8</th>
<th>2;9</th>
<th>2;10</th>
<th>2;11</th>
<th>3;0</th>
<th>3;1</th>
<th>3;2</th>
</tr>
</thead>
<tbody>
<tr>
<td>m &gt; f</td>
<td>100%</td>
<td>100%</td>
<td>88%</td>
<td>100%</td>
<td>80%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>n &gt; f</td>
<td>-</td>
<td>-</td>
<td>15%</td>
<td>-</td>
<td>40%</td>
<td>-</td>
<td>20%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>n=1</td>
<td>n=3</td>
<td>n=13</td>
<td>n=10</td>
<td>n=5</td>
<td>n=10</td>
<td>n=21</td>
<td>n=14</td>
<td>n=26</td>
<td>n=51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3;3</th>
<th>3;4</th>
<th>3;5</th>
<th>3;6</th>
<th>3;7</th>
<th>3;8</th>
<th>3;10</th>
<th>4;0</th>
<th>4;1</th>
</tr>
</thead>
<tbody>
<tr>
<td>m &gt; f</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>97%</td>
<td>100%</td>
<td>100%</td>
<td>98%</td>
<td>-</td>
<td>50%</td>
</tr>
<tr>
<td>n &gt; f</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3%</td>
<td>-</td>
<td>-</td>
<td>2%</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>n=7</td>
<td>n=26</td>
<td>n=11</td>
<td>n=38</td>
<td>n=3</td>
<td>n=1</td>
<td>n=65</td>
<td>n=1</td>
<td>n=2</td>
</tr>
</tbody>
</table>

Figure 11. Masculine as the replacing gender
Hania shows a strong preference for feminine as a replacement gender between 2;11-3;5. Neither before 2;11 nor after 3;5 is feminine realised as an alternative marker for other genders. As an explanation it could be suggested that between 2;11-3;5, Hania went through a stage of disambiguating feminine from masculine and neuter. Her overusing feminine marker -a changes into overusing masculine inanimate gender, which is used to replace masculine animate gender. If we leave aside three instances between 2;9-3;3,
masculine inanimate is mainly realised as a replacement gender between 3;6-4;1. Hania shows certain order: first she deals with perhaps the easiest marker -a, and then follows with animacy. If we contrast this mechanism with Jerzy’s database, there are hardly any feminine replacements, apart from with verbs. Understandably, it is very common for male children to copy feminine gender in past tense verbs from their mother’s input. Still, if we contrast Jerzy’s and Patrick’s use of feminine in past tense verbs we will see two very dissimilar tendencies: Jerzy only uses feminine gender in past tense verbs until the age of 3;10 (and possibly later), whereas Patrick never refers to himself using feminine past tense verbs. It is even more intriguing when we take into account the fact that Jerzy’s Polish input comes from one person (his mother), and Patrick’s from two (his mother and his older sister). Jerzy gradually learns to refer to himself using masculine past tense verbs, and whenever he does choose masculine, we know that he consciously makes a choice between an overwhelming amount of feminine marker in his mother’s input and the masculine gender. Singly occurring attempts are found at the ages of 2;8, 3;1, 3;2, 3;4, but from the age 3;6, the frequency of masculine occurrences on the 1st person past tense verbs increases gradually in Jerzy’s speech from 20% to 47% at the age 3;10.

Between 3;6-3;7 Hania makes more mistakes in semantics. On a few occasions, she refers to herself using masculine gender. What are the reasons? Masculine has been shown to be dominating after the feminine-sorting stage, and it is hypothesised that in the masculine-sorting stage it takes over not only non-semantic-based agreement cases, but also “obvious” cases such as a past tense verb referring to the girl. It supports the view that Hania is learning genders “one-at-a-time”, and to follow this principle she is prepared to compromise what she already knows is correct in order to accommodate learning the agreement rules for masculine after she has mastered those for feminine.

Jerzy’s database shows that in some months there are patches of feminine used instead of other genders. It is easy to fall into the trap and think that he went through a similar stage as Hania, but apart from a few adjectives and numerals between 2;5-3;7, the feminine gender in Jerzy’s data is mainly used with verbs in the past tense. In the last recording session at 3;10 he introduces feminine to more numerals and adjectives and uses feminine more often as a replacement in that session, which may suggest that a feminine sorting stage which Hania went through at the age of 2;11-3;5 begins in Jerzy’s speech at the age of 3;10. Undoubtedly, he must have acknowledged the feminine marker earlier, but reserved it for verbs, because it was also reinforced by his input. His mother would use feminine verb forms in the past tense, and Jerzy may have retained it for himself and treated it as correct. He invariably chooses feminine to refer to himself, even long after the recording sessions ended. In Jerzy’s case, it is difficult to distinguish what he believes to be correct.
Patrick went through a stage of when both masculine and feminine were used to replace other genders, unlike Hania, who tried to deal with feminine first, and overused it for a certain period before moving on to masculine.

CLASS 2
a. m (in)anim > m (in)anim

Animacy errors constitute about 13.5% of all gender errors. In early errors (around the age 2;5), masculine animate replaces masculine inanimate, and not vice versa. Judging by the proportions of the errors in the two groups, (see Table 23 group 2a), vice versa would be expected. Inanimate replacements\(^44\) for animate constitute 32% of all animacy errors, and it is only 8% in the opposite direction. The figures below illustrate both:

Figure 13. Masculine animate replaced by masculine inanimate

Examples:
(51) Znalazlem konik-a=konik-a\(^*\)  
find.1SG.PST.PFV.M horse(M.ANIM)-ACC.SG.DIM=M.ANIM.NOM.SG.DIM

morski-ego=morski-ego\(^*\)  
sea-M.ANIM.ACC.SG=M.ANIM.NOM.SG  
'I found a sea horse.'

(52) Taki-ego=taki-ego\(^*\)  
that-M.ANIM.ACC.SG=M.ANIM.NOM.SG  
piesek-a=piesek-a\(^*\)  
dog(M.ANIM)-ACC.SG.DIM=M.ANIM.NOM.SG.DIM  
'chcialem, mama.
want.1P.PST.M mum(F)
'I wanted that dog, mum.'

\(^44\) To avoid wordiness in describing animacy errors in Class 2, masculine animate and masculine inanimate nouns are sometimes referred to here as animate and inanimate nouns.
Figure 14. Masculine inanimate replaced by masculine animate

Examples:

(53)  
\[ \text{Ja zrobię straszającego} \]
\[ \text{I make-1P.FUT.PFV scary-M.ANIM.ACC.SG=MA.NAM.NOM.SG} \]

\[ \text{robaczka=robaczek-a} \]
\[ \text{bug(M.ANIM)-ACC.SG.DIM=MA.NAM.NOM.SG.DIM} \]

'I'll make a scary bug.'

(54)  
\[ \text{który-ego} \]
\[ \text{which-M.INAN.ACC.SG=MA.INAN.GEN.SG} \]
\[ \text{zje} \]
\[ \text{eat-3SG.FUT.PFV} \]

\[ \text{Pucuś jogurcik-a=jogurcik-ego} \]
\[ \text{Pucuś(M) yoghurt(M.INAN)-ACC.DIM=MA.INAN-GEN.SG.DIM} \]

'Which yoghurt will Pucuś eat?'

(55)  
\[ \text{przez ten-ego} \]
\[ \text{through this-M.INAN.ACC.SG=MA.INAN.GEN.SG} \]

\[ \text{tunel-a=tunel-ego} \]
\[ \text{tunnel(M.INAN)-ACC.SG=MA.INAN-GEN.SG} \]

'Through this tunnel.'

(56)  
\[ \text{mam takie specjalnie, żeby zepsuć} \]
\[ \text{have.3PL.PRES those-N_M.PERS on purpose to break.INF} \]

\[ \text{ten-ego} \]
\[ \text{this-M.INAN.ACC.SG=MA.INAN.GEN.SG} \]
\[ \text{zamek-ego} \]
\[ \text{castle(M.INAN)-ACC.SG} \]

'They have those on purpose to break this castle.'
Figure 15. Masculine animate-animate noun frequency (2;4-4;3)

<table>
<thead>
<tr>
<th></th>
<th>IN</th>
<th>AN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:4</td>
<td>42%</td>
<td>58%</td>
</tr>
<tr>
<td>2:5</td>
<td>65%</td>
<td>31%</td>
</tr>
<tr>
<td>2:6</td>
<td>51%</td>
<td>49%</td>
</tr>
<tr>
<td>2:7</td>
<td>57%</td>
<td>43%</td>
</tr>
<tr>
<td>2:8</td>
<td>58%</td>
<td>51%</td>
</tr>
<tr>
<td>2:9</td>
<td>50%</td>
<td>42%</td>
</tr>
<tr>
<td>2:10</td>
<td>67%</td>
<td>43%</td>
</tr>
<tr>
<td>2:11</td>
<td>77%</td>
<td>50%</td>
</tr>
<tr>
<td>3:0</td>
<td>65%</td>
<td>33%</td>
</tr>
<tr>
<td>3:1</td>
<td>59%</td>
<td>23%</td>
</tr>
<tr>
<td>3:2</td>
<td>71%</td>
<td>35%</td>
</tr>
<tr>
<td>3:3</td>
<td>85%</td>
<td>41%</td>
</tr>
<tr>
<td>3:4</td>
<td>82%</td>
<td>29%</td>
</tr>
<tr>
<td>3:5</td>
<td>61%</td>
<td>35%</td>
</tr>
<tr>
<td>3:6</td>
<td>54%</td>
<td>35%</td>
</tr>
<tr>
<td>3:7</td>
<td>60%</td>
<td>39%</td>
</tr>
<tr>
<td>3:8</td>
<td>55%</td>
<td>46%</td>
</tr>
<tr>
<td>3:9</td>
<td>64%</td>
<td>40%</td>
</tr>
<tr>
<td>3:10</td>
<td>58%</td>
<td>50%</td>
</tr>
<tr>
<td>3:11</td>
<td>65%</td>
<td>36%</td>
</tr>
<tr>
<td>4:0</td>
<td>61%</td>
<td>42%</td>
</tr>
<tr>
<td>4:1</td>
<td>35%</td>
<td>39%</td>
</tr>
<tr>
<td>4:2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table shows the frequency of masculine animate-animate nouns in the given age range.
Masculine animate nouns are present in the bilingual children’s speech from the beginning of the recording period (2;4). Although those early animate nouns are usually given the accusative case ending -a, in order to observe the animacy feature being acquired we need the evidence of agreement. According to the rules of Polish grammar, in an NP where the controller is a masculine animate noun, the accusative ending on the modifier as well as on the noun mirrors the endings in the genitive case. In an NP where the controller is an inanimate noun, the endings are the same in accusative and nominative. Let us consider ex. 51, which is an accusative utterance containing agreement between the masculine animate noun konik ‘horse’ and an adjective morski ‘sea’. The correct agreement endings for this animate noun and its modifier are konika morskiego. The child’s erroneous endings on the noun and the adjective reveal that by the age of 2;10 he had not learned the rules of animacy. Moreover, the same error type is found in his data a year later (ex. 53), when, instead of applying the animate endings -ego (adjective) and -a (noun) in his accusative utterance, the child uses nominative forms, which in accusative utterances are used with inanimate nouns. Children’s errors in case syncretism for masculine nouns (animate ACC=GEN, inanimate ACC=NOM) reveal their struggle with animacy, which they need to discover to follow these rules correctly and build adult-like utterances. Thus, any confusion involving animacy provides evidence that this feature requires more skill that just recognising whether a masculine noun is animate or not.

When discussing errors between masculine animate and inanimate, it is advisable to also consider the overall proportions of animate-inanimate nouns in the data. The children are expected to stretch the animacy rules of the more common subgender (m inanimate) onto the less frequent one (m animate). Apart from the animate-inanimate proportions, an additional factor must be considered here, namely the accusative case, through which the animacy rules are observed and learned. In the following paragraphs, the results of various frequency calculations carried out in the CLAN program are discussed.

Undoubtedly, the frequency of nouns of any gender depends on activities in which the children were involved during the recording sessions. The activities initiated by the investigator were often repeated in each home so that comparisons were possible as to the children’s linguistic competence at a given stage. Therefore, the children were bound to produce more animate nouns at sessions in which, for instance, an animal puzzle game was involved. This may serve as an explanation as to why at the age 2;6, 2;8, 2;11, 3;10 the frequency of animate-inanimate is nearly equal. The 40%-60% proportion of animate-
inanimate nouns found in the database (see Figure 15) and the 20%-80% proportion of animate-inanimate replacements (i.e. animacy errors) in agreement situations (see Table 23) are mapped in Figure 16. The left-hand axis is for the overall animate-inanimate frequencies, whereas the right-hand side one is for the animacy errors. The mapping of the noun frequencies and the animacy errors indicates two things: firstly, animacy errors are rare between 2;4-3;0, but they become more frequent after the age of 3;5; secondly, the frequency of animacy errors is not related to the general frequency of animate and inanimate nouns. Interestingly, there are no animacy errors whatsoever when the number of masculine nouns in the children speech is the highest at the age of 3;2. Two possible interpretations of the low amount of animacy errors despite the frequent use of masculine nouns can be offered: the children avoid agreement due to being in the process of figuring out other aspects of the gender system, or the children have acquired the rules of animacy very early on. The first interpretation falls in line with what has been reported with reference to the feminine gender in CLASS 1, namely that it is used very actively as a replacement gender between 3;0 and 3;7, which indicates that around that time the children are “busy” with the rules of feminine agreement.
Figure 16. Mapping of the overall animate-inanimate noun frequencies to animate-inanimate gender errors
The second interpretation, on the other hand, must take into account that after the age of 3;5 the children make many more animacy-based errors, which implies that the process of learning the animacy rules was not completed at an early age at all. This interpretation can be complemented in the following way: the children acquired the rules of animacy before 2;5, but around 3;0 they revisit the animacy-based rules, and overgeneralized masculine inanimate, producing erroneous agreement. The entire process presupposes that the children had acquired animacy rules before the recording sessions were launched (hence their initial animacy errors were missed), which would be sensible to think if examples of masculine agreement in the accusative case could be found in the earliest recordings, i.e. at 2;4 or 2;5. Instead, the database offers first examples of correct masculine agreement at the following stages:

<table>
<thead>
<tr>
<th>Inanimate</th>
<th>Hania</th>
<th>Patrick</th>
<th>Jerzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>3;2</td>
<td>2;7</td>
<td>2;9</td>
</tr>
<tr>
<td></td>
<td>2;11</td>
<td>2;11</td>
<td>3;10</td>
</tr>
</tbody>
</table>

A couple of animacy based errors can be found as early as 2;5, but correct agreement is not observed until much later: the children begin to notice and implement the rules of animacy around the age of 2;11-3;0. Compared to later months, the number of animacy errors between 3;0-3;5 is relatively low. This could result from the children avoiding decisions about animacy (and hence masculine agreement) until they have acquired the rules for the feminine agreement. To test this hypothesis, a thorough examination of the monolingual Polish data for ages 3;1-3;4 is required and it is presented below. The analysis includes all instances of masculine and feminine agreement, which are divided into two groups: correct and incorrect. Animacy has been taken into account for all masculine animate nouns and all those masculine inanimate nouns, which decline according to the masculine animate declension (e.g. *bajwan* ‘snowman(M)’, *lod* ‘ice-cream(M)*’). The main aim is to check the frequency of the correct and incorrect masculine and feminine agreement, as well as the proportions of animate-inanimate errors within masculine.

Overall, 62% of NPs with correct agreement found at 3;1-3;4 are masculine, and 38% are feminine. Incorrect agreement involves both genders nearly equally frequently. Both genders are definitely still “under construction” between 3;1-3;4, but such a high frequency of correct masculine agreement indicates that masculine is not “avoided” at that stage.
The question remains as to what percentage of the 20% of the incorrect masculine agreement included the animacy feature. This is illustrated in Figure 19:
Figure 19. Masculine correct-incorrect agreement in detail (3;1-3;4)

<table>
<thead>
<tr>
<th></th>
<th>3;1</th>
<th>3;2</th>
<th>3;3</th>
<th>3;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masc Incorrect</td>
<td>27%</td>
<td>16%</td>
<td>23%</td>
<td>17%</td>
</tr>
<tr>
<td>Masc Correct</td>
<td>73%</td>
<td>84%</td>
<td>77%</td>
<td>83%</td>
</tr>
<tr>
<td>n</td>
<td>85</td>
<td>148</td>
<td>58</td>
<td>23</td>
</tr>
</tbody>
</table>

Figure 20. Feminine correct-incorrect agreement in detail (3;1-3;4)

<table>
<thead>
<tr>
<th></th>
<th>3;1</th>
<th>3;2</th>
<th>3;3</th>
<th>3;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fem Incorrect</td>
<td>34%</td>
<td>26%</td>
<td>20%</td>
<td>15%</td>
</tr>
<tr>
<td>Fem Correct</td>
<td>66%</td>
<td>74%</td>
<td>80%</td>
<td>85%</td>
</tr>
<tr>
<td>n</td>
<td>67</td>
<td>107</td>
<td>44</td>
<td>13</td>
</tr>
</tbody>
</table>

Figure 21 compares the proportions of animate and inanimate agreement within the correct and incorrect masculine NPs. The vast majority of correct masculine NPs involves the inanimate gender (69%). Incorrect masculine NPs involve animate and inanimate nearly equally frequently. Figures 22 and 23 show a more detailed distribution of NPs including animacy at 3;1-3;4.
Figure 21. Overall correct-incorrect masculine animate-inanimate agreement at 3;1-3;4

Figure 22. Correct masculine animate-inanimate agreement in detail (3;1-3;4)
It is noticeable that the number of NPs with both correct and incorrect masculine animate agreement increased between 3;1-3;4, and by the age of 3;4, 70% of incorrect NPs that involve masculine gender are animacy-based (d.). The children may not be avoiding masculine between 3;1-3;4, but while the number of correct NPs involving feminine is growing (Figure 20), the number of incorrect animacy-based NPs is also increasing, which indicates the children’s growing readiness for and acceptance of animacy-based errors. This could be related to the rules of feminine agreement becoming clearer to the children, who, as a result, are feeling more capable of handling a new set of rules, i.e. animacy-based rules.

The second observation based on the mapping in Figure 16 is that the animacy error rates are not correlated with the frequency of animate and inanimate nouns. It is best shown at the ages 3;6, 3;8, 3;10, and 4;1, when the rates of inanimate-animate nouns and animacy-based errors show independent tendencies (Figure 24). Regardless of whether inanimate nouns are twice as frequent as (3;6) or only slightly more frequent than animate nouns (3;8), the same number of animacy errors is found. The largest amount of animacy-based errors occurs when both subgenders are equally frequent (3;10), which is either a result of a larger frequency of animate nouns involved in agreement, or a sign that by the age of 3;10 the children are more open to make animacy-based errors and finally learn the animacy rules. Patrick makes occasional animacy-based errors, which albeit rare, are continued until 4;3. Interestingly, his initial animacy-based errors are not ones in which he de-animates animate nouns. On the contrary – he applies animate endings to inanimate nouns.
The problem with transparency of the animacy rules is believed here to be related to the frequency of accusative in the children's speech, hence the next logical step is to check the frequency of nominative and accusative in the children's data. We now turn to this.

First, let us first briefly return to the animate-inanimate frequencies shown in Figure 15, which was generated with the use of the following CLAN commands:

a. freq +t tmpl +s*N:M* -s*PL:* -s*NPR* +d4
b. freq +t tmpl +s*N:AN* -s*PL:* -s*NPR* +d4

Based on the obtained total of masculine nouns in the singular (command a.), the number of only masculine animate nouns was calculated (command b.), and extracted from the total to

---

Dative has been omitted here due to lack of any data for this case.
receive the number of inanimate nouns. Since those results have already narrowed down the
search area to singular masculine nouns, the result files have been used to calculate the case
frequencies. Since CLAN requires test files to be in the .cha format, the output files from the
first analysis needed to be transferred to this format to be reliable in the case frequency tests.
Such a transfer can be easily done with CLAN through the textin command. Unfortunately,
textin does not preserve the number of tokens from the transcripts used for the initial
calculations. It produces a .cha file ready for further analysis, but saves only the number of
word types, rather than tokens (i.e. the amount of words types produced), yielding the output
of case frequencies (in fact, any information on morphological categories) unreliable. To
overcome this problem, original transcript files were used, and a more elaborate command
was built:

c. freq +tmpl +s*N:M* +s*SG: NOM* -s*PL:* -s*NPR* -s*SG:ACC* -s*SG:LOC*
-s*SG:DAT* -s*SG:INS* -s*SG:GEN* -s*SG:VOC* -s*N:F* -s*N:N*

For comparative purposes, below are two types of output for the SG:NOM frequency
for masculine gender. Output A. is done on a file converted by textin (which includes word
types only), whereas output B. is done on the child's original transcripts:

A. 666 Total number of different word types used
B. 1265 Total number of words (tokens)

It is crucial to be watchful for potential flaws in any research tool, and the above example
shows how to avoid obtaining unreliable results in CLAN.

The semantic rule plays an important role in allocating animate nouns to either
masculine or feminine. In the case of the contrast between masculine animate and inanimate
observable in the accusative case, the semantic rule is no longer important, as the choice in
agreement situation is between masculine inanimate or animate, not between masculine or
feminine. The frequency issue of animate and inanimate may therefore be an important
factor in explaining the overrepresentation of masculine inanimate agreement where animate
agreement is expected.

b. f/n > m (m)anim

In this group, the vast majority of the replaced genders involve feminine and 99% of
the replacing subgender is masculine inanimate. Figure 26 below presents the replaced
genders both in the accusative case (inanimate is distinguishable from animate), and in other
cases.
Figure 26. Feminine replaced by masculine in the accusative and non-accusative cases

<table>
<thead>
<tr>
<th>f &gt; m (non-Acc)</th>
<th>2;4</th>
<th>2;5</th>
<th>2;6</th>
<th>2;7</th>
<th>2;8</th>
<th>2;9</th>
<th>2;10</th>
<th>2;11</th>
<th>3;0</th>
<th>3;1</th>
<th>3;2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>94%</td>
<td>94%</td>
<td>70%</td>
<td>94%</td>
<td>60%</td>
<td>100%</td>
<td>94%</td>
<td>67%</td>
<td>68%</td>
</tr>
</tbody>
</table>

| f > m in (Acc)  | 6%  | 6%  | 6%  | 30% | 6%  | 40% | 100% | 94%  | 67% | 6%  | 33% | 32% |

n=1  n=4  n=17 n=16  n=10  n=18  n=5  n=21 n=16  n=27 n=26

Examples:

(57) *Miesz moj-q=moj-a* kolderka. [Jerzy 2;8]
    have.2SG.IMP my-F.ACC.SG=M.INAN.ACC.SG duvet(F)-ACC.SG
    'Have my duvet.'

(58) *On ma duz-q=duz-y* szyja. [Jerzy 3;10]
    he have.3SG.PRES large-F.ACC.SG=M.ANIM.ACC.SG neck(F)-ACC.SG
    'He has a large neck.'

(59) *Dlaczego Wendy ma drug-q=drug-i-ego? [Han 3;7]
    why Wendy have.3SG.PRES second-F.ACC.SG=M.ANIM.ACC.SG
    'Why does Wendy have a second hand?'

(60) *Przez cal-q-cale* noc* t-o. [Pat 4;3]
    through whole-F.ACC.SG=M.INAN.ACC.SG night(F).ACC.SG this-N
    'This playdough was open for the whole night.'
Perhaps the most defining characteristic in this group is the children's skilful use of the declension suffix preceded by the choice of the wrong gender on the agreeing modifier. In ex. 57, Jerzy makes a correct choice of the accusative ending -ę for the feminine noun kolderka 'duvet(F)', but an incorrect gender choice on the possessive pronoun môj 'my-M.ACC.SG'. Since the same process is repeated many times between the ages 2;8 and 3;10 (ex. 58), it can be suggested that the child has acquired (at least partly) the functional features related to case distinctions, but not the features related to gender distinctions. Other similar errors in this group involve masculine animate replacements, where the case distinctions are made correctly (rękę 'hand(F)-ACC.SG'), yet the gender distinction for the numeral drugiego 'second-M.ACC.SG' uncovers the child choice of masculine animate (ex. 59). Example 60 is different, as it involves a feminine noun ending in a consonant (noc), which does not take -ę in the accusative (like most feminine nouns ending in a vowel do). Two interpretations are possible: Patrick either inflects noc correctly, or treats it as masculine. The fact that noc is placed in agreement with caly-M indicates it is the latter.

Errors within plural

CLASS 3

(non)mp > (non)mp

Learning the masculine personal/non-masculine personal distinction requires additional effort from young speakers, since the plural genders in Polish are not a mirror reflection of genders in the singular, which contributes to the children's delay in learning correct agreement rules. The children must first recognise the fact that the plural agreement requires consistency in following a certain set of rules, similarly to the singular. However, they need to extend the rules acquired for the gender agreement in the singular to be able to observe and learn that the masculine personal group includes only nouns referring to men, and not all animate nouns (which is the case in the animate-inanimate subgender in the singular). They are required to pay careful attention to the animacy feature when learning the singular gender agreement, but in the plural, they are expected to recognise which noun represents a male person, and which does not, rather than just which nouns is animate and which one is not. The disproportion of the nonfeminine gender (masculine and neuter) versus feminine in the singular, and masculine personal versus non-masculine personal in the plural is partly mentioned when CLASS 1 errors are discussed. The similarities between masculine and neuter in the singular (expected to delay the children's learning of gender distinctions) may lead to neuter behaving like masculine rather than feminine in the plural.
Overall, the largest number of masculine personal/non-masculine personal errors is found after the age 3;4. As it is more common, non-masculine personal replaces masculine personal much more frequently. Below two examples are presented of how children confuse masculine personal and non-masculine personal pronouns when referring to one noun. Example 61 shows correct agreement between the plural noun *dinozaury* ('dinosaurs(N_M.PERS)') and the non-masculine personal demonstrative *te* 'these'. The correct status of this agreement results from the default character of *te* rather than Jerzy’s conscious choice of this non-masculine personal pronoun over the masculine personal *ci*. The interesting process takes place outside the NP, as the child refers to the same noun with the masculine personal personal pronoun *oni*. The reason why Jerzy uses *oni* is simple though: he overuses the pronoun in all recording sessions for this month. It is not a conscious choice, but an interplay of elements observed in the input, which he finally begins to implement. As earlier, his method is to overuse a form he chooses, before he is ready to move on to something new. The increased number of masculine personal replacements at the age 3;10 (Figure 27) is Jerzy’s overuse of the masculine personal pronoun *oni*, and not an example of the so-called “educated guess”.

The non-masculine personal NP *te klocki* ‘these-ACC blocks-ACC’ and the masculine personal pronoun *ich* ‘them-M.PERS’ in ex. 62 belong to two clauses separated in the original recording by five utterances. Similarly to Jerzy, Patrick arrives at correct agreement by using *te* as the default. Interestingly, when he is expected to produce an oblique form of the non-masculine personal pronoun *one* ‘they-N_M.PERS’ to agree with the controlling noun, he chooses the oblique form of the masculine personal pronoun *ich*, not the non-masculine personal form *je*. The answer seems to be hidden in the frequency of these pronouns in the adult Polish. Below in bold are marked all pronouns pertinent to this situation:

<table>
<thead>
<tr>
<th></th>
<th>'they'</th>
<th></th>
<th>'these'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M.PERSN_M.PERS</td>
<td></td>
<td>M.PERSN_M.PERS</td>
</tr>
<tr>
<td>NOM</td>
<td>oni</td>
<td>one</td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>ich</td>
<td>ich</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>im</td>
<td>im</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>ich</td>
<td>je</td>
<td></td>
</tr>
<tr>
<td>LOC</td>
<td>nich</td>
<td>nich</td>
<td></td>
</tr>
<tr>
<td>INS</td>
<td>nimi</td>
<td>nimi</td>
<td></td>
</tr>
</tbody>
</table>

The masculine personal pronoun *ich* occurs in the genitive and the accusative, and its derivative *nich* is also used in the locative. All these cases frequently occur in adult Polish. In contrast, the non-masculine personal *je* occurs only in accusative, and the masculine personal *ci* only in nominative. The explanation offered here as to why the children confuse
masculine personal and non-masculine personal when using pronouns is that it happens simply because of the varying frequency of correct forms which they receive in the input.

Figure 27. Masculine personal/non-masculine personal mutual replacements

![Graph showing mutual replacements between masculine and non-masculine personal pronouns]

<table>
<thead>
<tr>
<th>Masc pers &gt; Non-masc pers</th>
<th>Non-masc pers &gt; Masc pers</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% 100% 100% 100% 50%</td>
<td>n=5 n=7 n=1 n=4 n=4</td>
</tr>
<tr>
<td>83% 80% 100%</td>
<td>n=6 n=5 n=2</td>
</tr>
<tr>
<td>Masc pers &gt; Non-masc pers</td>
<td>Non-masc pers &gt; Masc pers</td>
</tr>
<tr>
<td>100% 95% 50% 100%</td>
<td>n=3 n=21 n=4 n=12</td>
</tr>
<tr>
<td>50% 67% 76% 100%</td>
<td>n=3 n=2</td>
</tr>
</tbody>
</table>

Examples:

(61) _Wiesz, jak robią te dinozaury_ i jak one=oni* chodzą?

'do know.2SG.PRES how do.3SG.PRES these(N_M.PERS)

'dinosaur(N_M.PERS).PL.NOM and how they(N_M.PERS)=M.PERS walk'

[Jerzy 3:10]

(62) MOT: _Samochód jest bardzo duży, Pucus._

'chodź?

'But it may destroy

'these building blocks.'

[Pat 3:1]
The examples below (63-66) show non-masculine personal as the replacement gender for masculine personal. Such an exchange is quite common in the children’s database and may be so because the non-masculine personal gender includes nouns from all three singular genders: feminine, neuter, and masculine. Interestingly, in ex. 65, the correct masculine personal nominative form *piraci* ‘pirates(M.PERS)’ (which requires alternation of the final consonant -i) constitutes an NP with the non-masculine personal indeterminate pronoun *jakieś* ‘some-N_M.PERS.NOM. Similarly, in ex. 66, the masculine personal noun *robotnicy* ‘workmen(M.PERS)’ is followed by the non-masculine personal form of the past tense verb *zrobify* ‘do.3PL.PST.N_M.PERS’. It suggests that the child has learned the features related to number or case distinctions, but not the features of gender agreement.

(63) \(\text{Ja tu buduję żąby ludzie} \)  
\(\text{I here build.1SG.PRES so that people(M.PERS).NOM} \)  
nie wypad-li^wypad-fy*.  
not fall.out-3PL.PST.M.PERS=3PL.PST.N_M.PERS  
‘I’m building here so that people don’t fall out.’

(64) \(\text{Ja lubię tak-ich=tak-le*} \)  
\(\text{I like.1SG.PRES such-M.PERS.ACC=N_M.PERS.ACC} \)  
\(\text{pan-ów=pan-y*} \)  
\(\text{co naprawiają} \)  
\(\text{man(M)-GEN.PL.M.PERS=GEN.PL.N_M.PERS who repair-3PL.PRES} \)  
dziurk-e.  
hole-F.SG.ACC  
‘I like these men who are mending the hole.’

(65) \(\text{Tu są jakieś=jakieś*} \)  
\(\text{here be.3PL.PRES some.M.PERS.NOM=N_M.PERS.NOM} \)  
\(\text{piraci}. \)  
\(\text{pirates(M.PERS).NOM} \)  
‘There are some pirates here.’

(66) \(\text{Robotnicy do-3} \)  
\(\text{workmen(M.PERS).NOM this-N.ACC.SG} \)  
zrobi-li=zrobi-ly*.  
do-3PL.PST.M.PERS=3PL.PST.N_M.PERS  
‘The workmen have done it.’

Some observations on the learning of masculine personal/non-masculine personal distinction by monolingual Polish children are offered by Smoczyńska (1985b), whose study is based on data between the ages 1;6 and 6;0, which is both much earlier and much later than the ages covered here. Smoczyńska reminds us that
...the criterion of grammatical gender distinction between feminine and neuter nouns on the one hand and masculine on the other is in two ways inconsistent from the point of view of the declension system. Firstly, in plural the distinction between masculine/feminine/neuter, relevant for singular, is abolished to be replaced with the virile/nonvirile^4^ gender distinction (1985b: 628).

Also, according to Smoczyńska, although such a process takes place in the adult language, "this conflict of double gender distinction does not actually occur because the masculine personal/non-masculine personal distinction with all its implications is acquired long after the appearance of the genitive plural forms" (Smoczyńska, 1985b: 628). It is unclear whether by "does not actually occur" she means that the reshuffling of the singular genders into masculine personal/non-masculine personal in the plural is not accounted for at all, or is present but not counted until correct masculine personal/non-masculine personal distinctions are properly acquired. The conflict itself seems to be a puzzle for the three children studied here since the moment they discover that both singular and plural nouns require some sort of agreement. Another unclear statement is that "agreement forms can also be of some importance for establishing gender distinctions" (Smoczyńska, 1985b: 629). It has been explained somewhere else in this thesis that agreement is the main tool for distinguishing the acquisition of the case distinctions from the acquisition of gender distinctions. Apart from syncretic cases, where agreeing controllers and targets take endings similar to one another, agreement is of vital importance in the establishing of the gender. How else do we observe gender distinctions between koty 'cats(N,M,PERS)' and kolesy 'colleagues(M,PERS)'? Both take -y as their plural marker, but once we begin counting them, we find ourselves distinguishing between the non-masculine personal cztery koty 'four cats' and the masculine personal czterej kolesy 'four colleagues'.

**Errors across number**

**CLASS 4**

(\textit{non})mp > m/f/n

This last error group includes errors in gender and number. Overall, the children seem to be making few errors involving both gender and number, although number errors as a separate group have been made quite frequently. The total amount of errors across number does not exceed 8% of all errors. Figure 28 shows the detailed distribution of singular gender replacements for non-masculine personal, and the main tendency is that neuter is slightly more frequently chosen than feminine. Such a trend is not altogether puzzling if we consider

^4^In this thesis, virile is referred to as masculine personal, and nonvirile and non-masculine personal.
the fact that feminine may actually be the first gender to provide children with some solid rules, whereas neuter on the contrary: the adjectival/numeral ending for neuter and non-masculine personal is identical (-e). It is puzzling, however, that although past tense verb forms for neuter singular and non-masculine personal are entirely different (-o and -y, respectively), the vast majority of neuter replacements for non-masculine personal take place on verbs. The syncretism of the -e ending on adjectives and numerals makes it impossible to establish if the child’s choice between neuter or non-masculine personal is correct, and thus, in all database utterances containing NPs with non-masculine personal nouns, adjectives and numerals are coded to match the noun. The child’s choice can only be observed in demonstratives (ex. 70) and verbs. Singular gender replacements for masculine personal are very scarce; hence only replacements for non-masculine personal are included in Figure 28.

Figure 28. Errors across number

<table>
<thead>
<tr>
<th></th>
<th>Non-masc pers &gt; m</th>
<th>Non-masc pers &gt; f</th>
<th>Non-masc pers &gt; n</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>310</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Examples:

(67) *Mama lubi duz-e=duz-y*  
`num(F) like.3SG.PRES big-N_M.PERS.ACC=M.INAN.ACC`  
rybk-i.  
fish(N_M.PERS)-ACC  
'Mum likes big fish.'

(68) *Nie, malutk-ie=malutk-i*  
`no tiny-N_M.PERS.NOM=F.NOM`  
gęsi=gęś*  
`geese(N_M.PERS).NOM=F.NOM`  
'No, the tiny geese.'

(69) *To jesz dla twoich=twój*  
`this be.3SG.PRES for your.GEN.PL=NOM.SG.M children(N_M.PERS).GEN`  
dzieci.  
'This is for your children.'
Examples 67-71 include nouns of various singular genders: feminine ending in -a (rybka 'fish(F)'), feminine ending in a consonant (geś 'goose(F)'), neuter (dziecko 'child(N)'), masculine animate (węż 'snake(M)') and masculine inanimate (cukiereczek 'sweetie(M)').

Treating them all as masculine even when they are in the plural form may be seen as a continuation of using masculine as the default gender in agreement situations, which Jerzy has also been found to employ in the mixed NPs (Analysis 4.2.2). However, it is striking that he recognises the plural number by using correct plural inflectional endings. In the examples above, he is unable to transfer the gender feature on the agreeing modifier, but in nearly all examples, the nouns are used in the correct plural form. It shows that Jerzy has acquired the correct declension, but gender distinctions are still a problem. This also shows why assigning declension on the basis of gender is likely to fail. We should ask whether the learning of number, case and gender at the same time delay learning of the last feature. It seems that all three features intertwine, but gender is the one causing most confusion. A couple of observations are required here. The noun dzieci 'children(N_M.PERS)' in ex. 69 is placed in an NP with the possessive pronoun twoj 'your-M' and is thus treated as masculine singular, despite being neuter in its singular form. It suggests that neuter and masculine may be treated as similar when used in the plural, just as they often are in the singular. In ex. 71, the plural form of the noun węż 'snake(M)' seems to be created by analogy to the preceding noun kobra 'cobras(N_M.PERS)'. Although kobra and węż both belong to the non-masculine personal gender, due to the varying quality of the final consonant they take different plural endings: kobra takes -y (it ends in a hard consonant), and węż takes -e (it ends in a soft consonant). It seems that it is quite a challenge for Jerzy to coordinate the requirements for agreement between the demonstrative pronoun te and kobra, and also produce two correct nominative plural forms based on the phonological information available to him. Only minutes away in the same recording, Jerzy produces correct and incorrect masculine personal agreement forms, which from the "half full glass perspective", suggests that around 3;10 he progresses from treating masculine as the absolute default to using other forms of agreement. Another
observation based on Jerzy's data concerns the use of neuter and masculine personal instead of non-masculine personal. Neuter replaces non-masculine personal between ages 2;4-3;5, and after 3;5 quite a clear shift to masculine personal is observable. Rather typically, masculine actively replaces non-masculine personal regardless of all other changes. The question arises: why does the shift in non-masculine personal genders take place so clearly between neuter and masculine personal? Jerzy does not use feminine as the replacing gender for any plural genders. Neuter is close to masculine in the singular, and Jerzy still makes mistakes when neuter is the correct gender by choosing masculine to replace it. We are left with masculine personal and the rules underlying plural genders as a place to look at. The shift seems to have little to do with learning of the "double gender distinctions", because masculine is still chosen in erroneous situations. Jerzy's attention is certainly switched from neuter to masculine personal when plural non-masculine personal agreement is required, and a chance of an explanation arrives from phonology of the personal pronouns. The shift is unlikely to be caused by the preference for an identical length of the masculine personal and non-masculine personal pronoun, because all three pronouns (neuter, masculine personal and non-masculine personal) have two syllables (ono, oni, one). Nevertheless, it could be suggested that it was caused by the phonological closeness of the masculine personal and non-masculine personal final vowels. The neuter final vowel is a back vowel (-o), whereas both, the masculine personal and the non-masculine personal (-i and -e respectively) vowels are front vowels. However, this explanation does not clarify the popularity of masculine replacements for masculine personal as far as pronouns are concerned (personal pronoun on, which is masculine, should be dropped in favour of the two-syllable oni, but it its not).

Hania's examples of the use of masculine personal as a replacement for non-masculine personal can be found quite early, i.e. at 2;11, compared to the other two children (Patrick 3;1, Jerzy 3;6). She is also the author of the only example of feminine singular replacement for masculine personal:

(72)  *Mama i tata nie wiedzieli-wiedzia-la*.

mum and dad not know-3PL.PST.M.PERS=3SG.PST.F

'Mum and dad didn't know.'

It is notable that although the obvious semantics of the father is the immediate reference for the past tense verb agreement, the replacement gender is feminine, and not masculine, which illustrates how the formal aspects of the noun tata are prioritised over the semantic aspects. The girl's example of the early use of masculine personal replacement is not followed with more masculine personal or non-masculine personal replacements until 3;5.
4.2.2 Mixed data

The area of language contact has provided researchers with intriguing, albeit contradictory data for many years. Language contact may manifest itself in numerous ways, among which are: interference, borrowings, transfer, code-switching, integration, among many others (Haugen, 1956, Romaine, 1995). Despite all the differences, these terms share one element: mixing of the linguistic material. This section provides more insight into the mixed utterances identified in the database with respect to grammatical gender. It needs to be explained at this point that by mixed utterances is meant the co-occurrence within one utterance of elements from both languages. Those elements can be of various types: phonological (traces of foreign sounds in words), morphological (e.g., inflectional suffixes), or syntactic (e.g., subject-drop in Polish). The primary reasoning employed here was that 2-3 year-old bilinguals would mix the two languages either as a result of incompetence, i.e. not knowing or not remembering proper terms in the language of the conversation, or as a strategy preventing a conversational breakdown. However, bilingual mixing found in the database may be more than that. Mixed utterances produced by Hania, Jerzy, and Patrick can be categorized from different perspectives. One can be that their insertions are simply “gap fillers” (see Sridhar and Sridhar, 1980). Singly-occurring insertions can also be seen as single-word code-mixing or lexical borrowing. To fully establish the status of the lexical material transferred between Polish and English in this database, an in-depth discussion regarding various perspectives on language contact phenomena would be necessary. This thesis does not aspire to cover all aspects of language contact, but a brief disambiguating discussion of the mixing mechanisms, constraints and theories is provided below. After that, the chapter will proceed with the analysis of mixed noun phrases.

The following terms will be defined in the forthcoming paragraphs: interference, transfer (production errors, simplification, developmental errors), borrowing and code-switching. As a precursor of this term, Weinreich (1970: 1) defines *interference* as those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact. This definition implies the rearrangement of patterns of one language resulting from the introduction of the foreign elements from the other language (compare Lehiste, 1988: 1). The contact of the languages takes place in the more highly structured domains of language, such as the phonemic system, a large part of morphology and syntax, and vocabulary. Romaine (1995: 60) points out that morphological interference is very common in bilinguals “by virtue of the fact that when speakers borrow words from one language to another, they may cause the morphology of the recipient language to be

---

49 Another term could be “nonce-borrowings” Poplack, S., Sankoff, D. & Miller, C. (1988).
realigned through the introduction of foreign morphemes. That type of interference may be found in Polish-English children’s speech, since Polish inflectional suffixes are numerous and can attach to English nouns easily. No attempts have been made to analyse all instances of interference in the present study. For a more detailed analysis, only utterances containing morpho-syntactic interference affecting grammatical gender, as well as utterances showing agreement between elements from the two languages in question, have been singled out.

Odlin (1989: 27) defines transfer as “influence resulting from similarities and differences between the target language and any other language that has been previously [...] acquired”, but admits that a “fully adequate definition of transfer seems unattainable without adequate definitions of many other terms” (1989: 28). This definition requires some extension to be fully satisfactory for the purposes of this study: transfer can also mean influence resulting from similarities and differences between two languages acquired simultaneously. Child language researchers, e.g., Müller (1998), have used the term “transfer” in their work to successfully describe instances of the influence between the bilingual child’s two first languages, and this practice will be employed in this study. A question needs to be addressed as to the distinction between interference and transfer. Weinreich’s definition of interference and Odlin’s of transfer overlap in saying that there is influence resulting from the contact between the two languages. According to Odlin (1989: 27), Weinreich used the term ‘interference’ to cover all cases of transfer, which after some reflection may seem a somewhat simplified approach. The variety of ways in which languages can influence each other has been captured by Odlin in his subdivision of transfer into: positive transfer, negative transfer and differing lengths of acquisition. Cross-linguistic similarities can result in positive transfer on many levels, but linguistically most intriguing cases result from negative transfer, which Lehiste defines as “deviations from the norms of either language that occur in the speech of bilinguals as a result of their familiarity with more than one language” (1988: 1). In that sense, Odlin’s negative transfer is very close to Weinreich’s notion of interference, the difference being that Odlin has introduced subtypes of negative transfer enabling a researcher to identify different cases of interference/negative transfer in the bilingual speech with greater detail.

Production errors (substitutions, calques and alterations) are a type of negative transfer. In L2 studies, substitutions occur when native language forms are used in the target language. As they are the easiest way to prevent communication breakdowns, they are the most common production errors found in the speech of bilingual children. In simultaneous bilingualism, the child creates substitutions by inserting forms from one language into an...
utterance otherwise in the other language. All examples below come from the author’s database:

(73)  *Król jest czerwony, a rycerz jest blue.*
    'The king is red and the knight is blue.'

(74)  *A ja chce pobawić się tymi animals, co trzeba wymówać.*
    'And I want to play with those animals which one has to take out.'

Calques are “errors that reflect very closely a native language structure” (Odlin, 1989: 37), which also involves structures copied from one language to the other. Calques frequently occur when the same grammatical structure in less complex in one of the languages, and an example here comes from Jerzy, who as a Polish-English bilingual, uses the English dative construction “to X” in a Polish utterance before he is able to produce a correct Polish dative inflectional suffixes:

(75)  INV: *Komu dziewczyn-ka daj-e kwiatki?*  
    who.DAT girl(F)-DIM give-PRES.IMPF.3SG flowers(N_M.PERS)  
    ‘Who is the girl giving the flowers to?’

CPL:  *Do dziadek*.
    to grandpa(M).NOM.SG  
    ‘To grandpa.’

Jerzy is “falling back” on the rule he knows in one of his languages, and is adapting or copying it to the other language to convey information the investigator is asking about. This mechanism produces an error, since the English rule does not correspond directly to the Polish dative structure, but as an easy and productive mechanism preventing a communication breakdown, it is frequent in bilingual speech. Calques will not be discussed in this thesis, since they are rarely (or never) related to gender errors.

Selinker’s (1972) concept of interlanguage originally denoted a unique intermediate system that emerges during the process of L2 acquisition and that differs from L1. Interlanguage is not a hybrid language consisting of elements from both languages. A learner “proceeds through a series of interlanguages on the way to complete mastery of the target language” and these interlanguages are more like an “intermediate system characterized by features resulting from language-learning strategies” (Appel and Muysken, 2006: 83). When “the learner postulates a simpler structure in his or her interlanguage than the one truly characterizing the target language”, an error of simplification occurs (Appel and Muysken, 2006: 90). In other words, a speaker may omit various parts of speech in one language due to their absence in the other, regularize structures such as word order, or use simplified
subordinate terms rather than contextually more fitting terms. A bilingual child's interlanguage may be full of simplifications on various levels, which may be a result of the doubled amount of vocabulary items and syntactic structures to learn. Moreover, the distance between the child's two languages may be affecting the process of acquisition by generating simplifications and other types of transfer errors. It is very possible that a similar mechanism applies in the BFLA (bilingual first language acquisition). The greater the structural difference between the same features in the languages involved, the greater the learning problem may be.

The last important type of negative transfer is developmental errors, which constitute developmental processes of identical nature, occurring both in monolingual and bilingual children. For instance, the overproduction of the genitive inflectional suffix -ów, which is mentioned by Smoczyńska (1985b: 628), is also observed in the author's bilingual database. Detecting them may need a control monolingual group for comparative purposes. It is of utmost importance to distinguish developmental errors from production errors in order to validate results based on the available data. The initial analysis of the mixed utterances recorded for this project has shown many errors evidently linked to the fact that the languages involved have been in contact. Apart from calques and substitutions, there are numerous examples of structurally conjoined lexical items, containing elements from both languages. In the example below, Hania mixes the verb to sleep with the Polish infinitive ending, creating a hybrid verb slipiać. Such examples should be treated as production errors.

(76)  Nie, Jilly musi **slipiać**.
'No, Jilly has to sleep.' [Haa 3:6]

Haugen defines borrowing as "the attempted reproduction in one language of patterns previously found in another" (1950). In the lexicon, borrowed words can be for example content words. In structure, borrowings may involve minor structural changes (e.g. derivational morphology), moderate structural changes (e.g., inflectional morphology), and heavy structural changes (e.g. leading to typological change). As has been mentioned earlier, interference is believed to result from the incompetence in one (or both) of the languages, and leads to the inadvertent mixing of the languages. Code-switching, however, is believed to be intentional. Also, contrary to interference (aimed at complete integration of the elements from a language with the syntactic/morphological rules of the recipient language), code-switching is based on the maintenance of the original form of words, phrases and sentences used by code-switchers (Hagège 1996: 239). Although words, phrases and sentences in code-switching remain "under control" of the source language, the users are able to identify the language and understand the content of the message sent by the
interlocutor. In other words, code-switching is usually intentional, and as such, it is often used to refer to more mature bilingual speakers. Children code-switch intentionally as well, but not until they reach a certain stage in their linguistic awareness. We now turn to the analysis of the mixed utterances, or more specifically, mixed noun phrases.

The database used for this project totals 56,501 children’s utterances, of which 1,322 are mixed Polish-English utterances containing mixed noun phrases (NPs). In these NPs, Polish parts of speech (PoS), e.g. adjectives, pronouns, numerals (attributive), as well as the 3rd person verbs in the past (predicate) are “in agreement” with English head nouns. The term agreement is used here to refer to the relation between English nouns and Polish PoS marked for gender, in this case a past tense verb form. Another place for Polish gender to manifest itself in mixed utterances are inflectional suffixes attached to the English nouns. Unfortunately, although a few nouns have received such suffixes, and a few are declined, there is no evidence of agreement, which does not allow us to identify the gender class with absolute certainty. In the total of 226 utterances with potential examples of gender assignment, 207 English nouns are allocated to one of the gender classes through agreement with one (or more) PoS, whereas 25 nouns have acquired inflectional suffixes, potentially placing them in one of the four Polish declensional classes. Such placement, however, does not prove that the noun has been allocated to one of the gender classes (to show this agreement is required for these 25 insertions). In a few instances, two modifiers belonging to different gender classes are used with one noun, thus exhibiting different gender agreement in the same utterance.

Polish and English NPs can be used in similar syntactic positions, and they express similar pragmatic functions. In other words, they are quite similar on the functional level. However, although the internal structure of singular NPs and the PoS that can be used within them are similar, there are major differences in the paradigmatic choices that need to be made: in English, a natural gender rule operates predominantly on personal and possessive pronouns, but not on adjectives, verbs, demonstrative pronouns etc. In Polish, on the other hand, gender is a morpho-syntactic category, and it largely determines the form of other PoS, in combination with a natural gender rule. In order to speak approximately like most adults around them and produce formally acceptable NPs, one of the main tasks that the three Polish-English bilingual children are faced with is marking of syntactic and natural gender in Polish and marking of just natural gender in English. The Polish gender system is a very pervasive category present on many modifiers. Therefore, the children are faced with decisions concerning gender assignment each time they insert an English noun and use it in

---

31 The term agreement has also been used with reference German-English mixed NPs by Endesfelder Quick (2008).
an agreement situation. They cannot escape the fact that as soon as they insert such a noun, the Polish morphosyntactic frame surrounding that noun has to somehow respond to that insertion. Examples of various PoS marked for gender within mixed NPs will show that the children are attempting to allocate the inserted nouns to one of the Polish gender classes by choosing the gender of the PoS appearing with that noun in the mixed NP. Their attempts have various outcomes, and at times are far from definitive. Nonetheless, the data provides evidence that the children's morpho-syntax does react to the insertions similarly to the way it would react to a Polish noun inserted in those NPs. The questions to ask here are: Is any of the genders preferred? Does that preference change with time? Which PoS are marked for gender most often in mixed NPs? Answers to these questions will provide the basis for comparison between the mixed data and the monolingual Polish data, and will help us decide whether Polish-English children use language specific or non-language specific patterns.

A word needs adding with reference to Polish equivalents of the English insertions. A noun in a monolingual Polish or monolingual English utterance has only one semantic level: its referential meaning. However, once an English noun has been inserted into a Polish utterance (or vice versa), another level emerges, i.e. the translational level. *Dog* inserted into a Polish utterance now also means *pies*. This additional semantic level opens the possibility for the fact that the Polish equivalents may play a role in guiding the child in choosing the gender for the English nouns. In their analysis of gender agreement in German-English mixed NPs, Endesfelder Quick at al (2008) use the terms correct and incorrect with reference to mixed NPs where the gender of a given insertion is the same as or different from the translation equivalent. These terms are inaccurate, since gender assignment involving English nouns (which lack grammatical gender) is hardly a prescriptive matter. To justify the use of terms correct/incorrect we would have to work within the hypothesis that the children refer to the translation equivalents each time they borrow an English noun. Ought we not to allow for the fact that at times they do not use semantic clues at all? If we do, we cannot term their choice as either correct or incorrect. The terminological problem described here has a bearing on the entire §4.2, in which many comparisons between the mixed and the monolingual Polish data is made. Therefore, when discussing the gender choices made by the children with reference to English insertions, instead of terms correct/incorrect, we choose terms matching/non-matching.

**Analysis 1 Mixed data**

A clear tendency is observable for all the children for a low number of mixed NPs between 2;4-3;4, and for an increased frequency between 3;6-3;11 for two of the three
children (Figure 29). The increased production of mixed NPs at certain age points may suggest that those two children went through a more active gender-marking phase, at least in the mixed utterances. Such activation could result from increased attempts to deal with English nouns creeping into their Polish structures, and consequently, attempting to understand the rules for gender-marking and gender agreement in Polish. The increase in the number of mixed NPs shows that the children were possibly more adventurous, or more confident, about the marking of gender on the English nouns. Mixing of the two languages in general, i.e. not only of NPs, is shown in Figure 30.

Figure 29. The frequency of mixed NPs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hania</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>33%</td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Patrick</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Jerzy</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>67%</td>
<td>64%</td>
<td>-</td>
<td>-</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hania</td>
<td>22%</td>
<td>-</td>
<td>14%</td>
<td>19%</td>
<td>14%</td>
<td>14%</td>
<td>14%</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Patrick</td>
<td>76%</td>
<td>-</td>
<td>29%</td>
<td>28%</td>
<td>76%</td>
<td>-</td>
<td>92%</td>
<td>92%</td>
<td>-</td>
<td>-</td>
<td>67%</td>
</tr>
<tr>
<td>Jerzy</td>
<td>100%</td>
<td>57%</td>
<td>56%</td>
<td>9%</td>
<td>8%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
<td>100%</td>
<td>-</td>
</tr>
</tbody>
</table>

The discrepancy between the onset of general mixing for both Hania and Patrick in comparison to Jerzy is striking mostly due to the fact that Jerzy’s mixed utterances appeared 6 months before the other two children. Moreover, the frequency of his mixing is highest in the first 6 months, contrary to the pattern observed for Hania and Patrick. Their mixing frequency increases with time, and apart from some months where the frequency is zero due to the missing data, it continues to increase with time. Jerzy represents a different mixing pattern to Hania and Patrick. Another observation refers to the fact that at the age of 3:0 and 3:6 the mixing in the speech of all three children is on the more or less the same level.
interesting that in between those two points the children's mixing was either low or medium high, with no extreme points. The children's individual mixing frequencies (with regard to mixed utterances in general, and mixed NPs in particular) is presented below in the charts, which make the discrepancy between Jerzy’s mixing and Hania and Patrick’s even clearer. The trendlines for Hania and Patrick show a gradual increase in their overall mixing rates. Hania’s mixed NPs increase only fractionally between the onset and the final recording, whereas the overall mixing increases significantly. In Patrick’s case, the situation is quite similar but for a marginally bigger increase of mixed NPs. The trendlines for Jerzy, on the other hand, point to a gradual decrease of the general mixing rates, and a very slow fall of the number of mixed NPs. Further questions may be asked: why is the rate of mixed NPs so low in comparison to the overall mixing rate? What influences the onset of the children’s mixing? What are the factors that condition Jerzy’s mixing, which is so different the mixing of the other two children? These and other questions require an additional study focused solely on the nature of the mixing phenomenon.

Figure 30. The frequency of mixed utterances

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hania</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4%</td>
<td>17%</td>
<td>11%</td>
<td>5%</td>
<td>24%</td>
<td>13%</td>
<td>14%</td>
<td>5%</td>
</tr>
<tr>
<td>Patrick</td>
<td>-</td>
<td>7%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8%</td>
<td>27%</td>
<td>53%</td>
<td>67%</td>
<td>29%</td>
</tr>
<tr>
<td>Jerzy</td>
<td>100%</td>
<td>93%</td>
<td>100%</td>
<td>96%</td>
<td>83%</td>
<td>89%</td>
<td>86%</td>
<td>49%</td>
<td>34%</td>
<td>19%</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td>n=6</td>
<td>n=73</td>
<td>n=80</td>
<td>n=70</td>
<td>n=42</td>
<td>n=47</td>
<td>n=83</td>
<td>n=85</td>
<td>n=32</td>
<td>n=36</td>
<td>n=76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hania</td>
<td>45%</td>
<td>13%</td>
<td>52%</td>
<td>33%</td>
<td>72%</td>
<td>100%</td>
<td>100%</td>
<td>2%</td>
<td>100%</td>
<td>100%</td>
<td>25%</td>
<td>-</td>
</tr>
<tr>
<td>Patrick</td>
<td>49%</td>
<td>6%</td>
<td>33%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>56%</td>
<td>-</td>
<td>-</td>
<td>75%</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>Jerzy</td>
<td>6%</td>
<td>87%</td>
<td>42%</td>
<td>34%</td>
<td>28%</td>
<td>-</td>
<td>-</td>
<td>43%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>n=71</td>
<td>n=36</td>
<td>n=86</td>
<td>n=150</td>
<td>n=88</td>
<td>n=83</td>
<td>n=2</td>
<td>n=122</td>
<td>n=5</td>
<td>n=5</td>
<td>n=36</td>
<td>n=13</td>
</tr>
</tbody>
</table>
Figure 31. Hania’s mixing rates

Figure 32. Patrick’s mixing rates

Figure 33. Jerzy’s mixing rates
Analysis 2 Various tendencies

Agreement involving English insertions may uncover interesting details about the ways bilingual Polish-English children discover and use the rules underlying the gender systems in both languages. Analysis 2 provides answers to the questions asked earlier: Is any of the genders preferred? Does that preference change with time? Which PoS are marked for gender most often in mixed NPs? It has already been established that, on some basis, the English insertions are allocated to one of the Polish gender classes, and Figure 34 shows how genders are distributed in mixed NPs. The results are based on three types of mixed NPs: one/two PoS + a singular English insertion (e.g. *u nas byl-o t-en fox* 'at ours was-MASC that-MASC fox(M)'); two PoS not in agreement with each other + a singular insertion (e.g. *zbudujemy tak-i moj-q road* 'we’ll build that-MASC my-FEM road(F)'); a singular PoS + plural English insertions (e.g., *drugi lights* 'second-MASC lights(N_M.PERS)'). The last type is included on the grounds that this distribution is to show the children’s gender choices in the singular, and plural insertions do not alter the fact that the children have chosen a particular gender for the PoS used.

As is clear from Figure 34, the children used PoS marked for masculine gender most often, thus exhibiting a strong bias towards it. Only a handful of nouns have been allocated to the non-masculine gender. There are only two cross-marking cases where both masculine and feminine gender is marked on two PoS used with reference to the same English insertion. There is only one such case where masculine and neuter gender is used with reference to the same insertion.

**Figure 34. Genders in mixed NPs**
It is important that we try to shed more light on the children’s bias towards masculine. In her study on English loanwords in contemporary Polish, Mańczak-Wolfheld (1995) found out that around 76% are allocated to masculine gender, and only about 24% are either non-masculine or unallocated. We examined the bias towards the masculine gender in this inventory, and found out such biases are strong; thus, the question about gender assignment in mixed NPs is not about masculine vs. feminine vs. neuter, but masculine vs. non-masculine. The strong bias towards masculine may suggest that the children are treating English nouns in a similar way to the way adult Polish speakers treat English loanwords. To test the bias, a $\chi^2$ test has been used. No significant difference between the children has been found as to how often each child chooses the corresponding Polish gender for words that are masculine in Polish, i.e., they are equally often “right” in choosing masculine. Taking into consideration that masculine is the most frequent gender for loanwords in Polish, it is expected that the children’s choice is more often matching in choosing masculine than feminine or neuter. Another question is: how often does each child choose the corresponding Polish gender for words that are feminine or neuter? Again, no significant difference between the children has been found, i.e., they are equally often “wrong” in choosing feminine or neuter. The masculine is favoured in the gender assignment of the mixed utterances.

The distribution in Figure 34 shows only a fractional amount of non-masculine gender assignment in the mixed NPs. On that basis we hypothesized that the bias towards masculine gender is shared in the three children. A closer look at the data shows that the over-representation of the masculine in the mixed NPs has its main source in Jerzy’s data, whereas for the other two children, it is not entirely clear whether their preference comes from the corresponding Polish gender or random gender assignment. It is surprising, however, that although Jerzy over-represents masculine, and as a result seems to be “performing better” in the mixed utterances than the Polish utterances, he seems to be the
most consistent one out of all three children. Hania and Patrick are clearly more often in doubt as to the gender assignment in the mixed utterances, despite their very good gender error results in the monolingual Polish data. For Jerzy, marking of masculine in the mixed NPs is in a way simply an extension of the rules applied to the Polish nouns. This is the first indication that there exists a division between the children: Hania and Patrick follow one path, and Jerzy follows another. In fact, it is even more than that: the three children represent two very different methods of assigning English insertions to Polish gender classes: Hania and Patrick treat the insertions in a very different way than they do Polish nouns. Any clues or rules applied to Polish nouns seem to disappear when they insert an English noun. Jerzy’s approach is different: English nouns do not receive special treatment; they are allocated to masculine just as many Polish nouns are. His overrepresentation of masculine confirms his treatment of masculine as the default for both Polish and English nouns.

Let us now analyse a series of comparative charts presenting children’s gender choices in mixed and Polish NPs. The dominance of masculine gender in mixed NPs indicated earlier is also observed in the children’s individual history of gender choices between the first and the last mixed NP in their database. In terms of mixed NPs, Hania (Figure 36) and Patrick (Figure 37) are similar – little marking of non-masculine gender all along, with masculine marked steadily at the beginning, but sharply increasing between the age 3;6-4;1. We can observe a clear correlation between the high number of masculine gender in mixed NPs between 3;6-3;11 and the high number of occurrence of masculine as a replacement gender for feminine and neuter in the monolingual database between 3;2-3;10. Taking into consideration the high level of feminine at the age 2;11-3;6, Hania shows a clear one-gender-at-a-time strategy. She makes a special effort to figure out feminine first before she moves on to the masculine (3;4-3;10).

Jerzy (Figure 38) is similar to the other two children in terms of poor non-masculine marking, but his marking of masculine differs significantly in that it increases at three different points. This confirms what has been observed earlier: at least with regard to the mixed NPs, the children form two groups: Hania and Patrick follow a similar route, whereas Jerzy follows a different route. Although Hania and Patrick’s approach to gender marking in the monolingual data is not as close to each other as it is in the mixed data, their Polish charts (Figures 39 and 40) show a similar tendency: all three genders are all actively marked. Jerzy’s chart (Figure 41) shows minimal marking for feminine and neuter, and strong “over-marking” of masculine. Jerzy clearly prefers masculine both in the mixed and the monolingual inventory. Hania and Patrick show such a preference in the mixed data, whereas in the monolingual Polish data they make use of all the three genders.
Figure 36. Hania's gender preference in mixed NPs

Figure 37. Patrick's gender preference in mixed NPs

Figure 38. Jerzy's gender preference in mixed NPs
Figure 39. Hania’s gender preference in Polish NPs

Figure 40. Patrick’s gender preference in Polish NPs

Figure 41. Jerzy’s gender preference in Polish NPs
The fact that all three children prefer masculine suggests that it is to a certain extent the "safe" option for them, at least in the mixed utterances. However, masculine is not a default choice in the monolingual utterances in the same way as it is in the mixed utterances. In the mixed NPs, masculine is the default for all three children, but in the monolingual utterances, it is the default only for Jerzy. Hania's gender choices in Figure 36 show that she is the most balanced in the group in terms of gender marking. Patrick shows a slight preference for masculine, but the frequency of feminine in his Polish data is also high in relation to masculine, compared to Jerzy's database, who is particularly biased towards masculine. Neuter is the least frequently marked gender, but its frequency in Hania and Patrick's data is rising over time, while in Jerzy's data it is falling.

The numbers for the monolingual Polish data presented in Figures 39-41 provide trendlines pointing to three different tendencies: 1) in Hania's data, on the whole, masculine and feminine receive equal amount of attention, but there is a gradual increase in preference for feminine agreement than masculine towards the age 4; 2) in Patrick's data, the preference for masculine agreement is significant and increasing, but feminine and neuter agreement is noticeable and also increase; 3) in Jerzy's data, the preference for masculine is strong all along; neuter agreement decreases with time, replaced by increasing feminine.

We now turn to the distribution of parts of speech (PoS) in mixed NPs (Figure 44). Demonstrative pronouns are involved in 47% of cases. Adjectives come next on the frequency scale, and they constitute around 1/4 of total number of PoS involved. Possessive pronouns and numerals constitute around 8%, whereas verbs 6%. Only 3% of agreement cases involve indeterminate and interrogative pronouns, and only 1% of relative and personal pronouns were involved in agreement situations in the mixed NPs. Noticeably, demonstrative pronouns and adjectives are most numerous, which points to the preference for the attributive position. A low frequency of verbs occurring in agreement with English nouns shows that the predicate position is less popular than the attributive position. Based on the frequency of PoS in mixed NPs, the following order begins to emerge:

dem pro > adj > num+poss pro > verb > indet pro+rel pro > inter pro+pers pro

The frequency of occurrence (in agreement situations) falls as we move from left to right. A strong tendency is observed for demonstrative pronouns to take part in agreement with

52 For comparative purposes verbs have also been included in the charts, although they do not belong to a noun phrase.
inserted nouns. Yet, it is likely that a different language pair would demonstrate a tendency quite different from the one observed here. Nevertheless, despite the fact that the children's individual frequencies are different (Figure 44), they all show a falling tendency as we move from the attributive position (demonstratives, adjectives, possessives) towards the predicate (verbs) and further down towards relative pronouns. This order may alternate even among speakers of the same language pair: Patrick's preference for demonstrative pronouns is not shared by Jerzy and Hania, who tend to modify their English insertions using adjectives mostly (Jerzy) or adjectives and demonstrative pronouns with similar frequency (Hania). Nonetheless, all three of them express a preference for attributive position (demonstratives, adjectives, numerals, possessives) over predicate (verbs) and relative pronouns.

The strong bias towards masculine (Figure 34) and the overwhelming preference for PoS in the attributive position in mixed NPs (Figure 44) provides us with a good basis for a comparison between the mixed data and the monolingual Polish data. The question here is: Do the frequencies of genders and PoS in mixed NPs correspond to those used in agreement situations in correct\(^3\) Polish NPs? If the preferences overlap, it will suggest that the children may be using non-language specific mechanisms when they are dealing with gender assignment of both Polish and English nouns within the frame of the Polish morpho-syntax. However, if the preferences differ significantly, it will suggest that they produce agreement situations according to more language-specific patterns. Two transcripts from different months per child have been picked from the Polish inventory according to two criteria: first, the child's mixing rate had to be high that month, and second, the child had to produce mixed NPs in which the gender of an insertion was both the same as and different from the Polish equivalent.

The comparative chart (Figure 43) shows a similar tendency both in Polish NPs and mixed NPs, namely that demonstrative pronouns and adjectives are most common, whereas different types of pronouns are least common. The preferences for PoS in both inventories are similar: falling towards relative and personal pronouns. The overlapping preference of modifiers in the attributive position in both mixed and Polish correct NPs suggests that in gender assignment of Polish and English nouns within the Polish morpho-syntax the children are using non-language specific mechanisms.

\(^3\) A correct Polish NP involves correct agreement; an incorrect NP involves erroneous agreement. Erroneous NPs are discussed in detail in 4.2.1.
The above comparison involves mixed NPs and Polish NPs with correct agreement. Let us also compare mixed NP and erroneous Polish NPs (discussed in greater detail in §4.2.1.). In erroneous Polish NP, the three most frequent PoS are verbs (31%), demonstrative pronouns
(27%) and adjectives (21%). The remaining PoS (possessive pronouns, numerals, personal pronouns, indeterminate and interrogative pronouns) do not exceed 7%. Demonstrative pronouns and adjectives taken together are most numerous (48%), which points to the preference for the attributive position. However, contrary to mixed NPs, the high frequency of verbs in erroneous Polish NPs shows that the predicate position is popular in the Polish data as well. Both situations are compared in Figures 44-45:

PoS in mixed NPs

Figure 44. PoS in mixed NPs

<table>
<thead>
<tr>
<th>PoS</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dem Pro</td>
<td>48%</td>
</tr>
<tr>
<td>Adj</td>
<td>24%</td>
</tr>
<tr>
<td>Num</td>
<td>8%</td>
</tr>
<tr>
<td>Poss Pro</td>
<td>8%</td>
</tr>
<tr>
<td>Vb</td>
<td>11%</td>
</tr>
<tr>
<td>Indet Pro</td>
<td>3%</td>
</tr>
<tr>
<td>Rel Pro</td>
<td>2%</td>
</tr>
<tr>
<td>Inter Pro</td>
<td>1%</td>
</tr>
<tr>
<td>Pers Pro</td>
<td>1%</td>
</tr>
</tbody>
</table>

n=194

Figure 45. PoS in Polish incorrect NPs

<table>
<thead>
<tr>
<th>PoS</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vb</td>
<td>323%</td>
</tr>
<tr>
<td>Dem Pro</td>
<td>278%</td>
</tr>
<tr>
<td>Adj</td>
<td>224%</td>
</tr>
<tr>
<td>Poss Pro</td>
<td>69%</td>
</tr>
<tr>
<td>Num</td>
<td>66%</td>
</tr>
<tr>
<td>Pers Pro</td>
<td>53%</td>
</tr>
<tr>
<td>Indet Pro</td>
<td>30%</td>
</tr>
<tr>
<td>Inter Pro</td>
<td>7%</td>
</tr>
</tbody>
</table>

n=1050
The differing preferences for PoS showed in the figures above weaken the claim that the children use non-language specific mechanisms. In addition, the children have different gender preferences in mixed NPs and in erroneous Polish NPs. The overwhelming bias towards masculine is identical in both contexts (Figure 34), but the use of feminine differs significantly: mixed NPs involve only 8% of feminine, whereas erroneous Polish NPs involve feminine in 34%. This favours the idea of the children using language-specific methods.

Analysis 3 Which clues?

This section discusses potential reasons guiding the children in allocating the inserted English nouns to a particular gender. Phonological factors are presented first, and are followed with a discussion on parts of speech used in mixed NPs. Finally, a section is devoted to animacy.

As we are considering phonological factors, the children’s English input requires some attention here. As has been described in §3.2, the children’s fathers are native speakers of English, and each speaks with a different accent. Hania’s father speaks with a mixture of London accent and Standard British RP, Jerzy’s father speaks with the Standard RP, whereas Patrick’s father speaks with a New Zealand accent. Children receive additional input in English from TV programmes and films, in (the majority of) which Standard English RP is used. The question to ask here is whether the different accents spoken by the fathers (as the primary sources of English input) have any bearing on the classification of the borrowed English nouns into the three groups described above. The answer is no. The fathers’ accents would have a direct impact here if they pronounced nouns such as monster or fire (spelled with the final -er/-re) with the rhotic -r, which would allocate them to Group I (nouns with consonantal endings). The fathers pronounce the final -r in nouns spelled with -er/-re as [ə] sound (schwa), placing them in Group II (non-consonantal endings). Although the fathers tend not to pronounce the final -t (their speech involves glottal stops), in their conversations their speech is slower and more careful, which renders their glottal stops far less frequent when speaking to the children than when addressing adults. Pronounced or not, however, as a consonantal sound, glottal stop does not have a bearing on the placement of nouns ending in -t, e.g. hat or teapot.

The data to be examined here have been classified into three groups:

\[\text{IPA}^{54}\]

\[\text{stressed [\text{\v{a}}]} \text{ or unstressed [\text{\v{a}}], as represented in the IPA}\]
- Group I - nouns with consonantal endings allocated to masculine
- Group II - nouns allocated to masculine but not ending in a consonant
- Group III - nouns with consonantal endings but allocated to a non-masculine gender

Group I contains nouns ending in one of the following consonantal codas\textsuperscript{55}: -n, -s, -t, -k, -g, -d, -m, -p, -w, -l, -v, -z, -lj, -dz, -η. Group II consists mostly of nouns that end in [a] or [i]/[ii]. A few nouns also end a diphthong, for example [ta], [di] or [zu]. Two isolated examples end in a long vowel, [di] and [zui]. With the exception of one noun ending in [zi], Group III nouns end in one of the following consonants: -n, -s, -t, -d, -g, -n, -p, -l. Group I is undoubtedly the largest. Since consonants characterize masculine nouns in Polish, mere observation of the data suggests that the dominance of the masculine gender in the children's gender choice for the English insertions is mostly due to the consonantal endings of the English nouns. With the exception of the nasal velar sound [-ŋ], all the codas included in Group I are present in both the Polish and the English phonetic repertoire. On this basis it can be claimed that the children are guided by the consonantal endings of the borrowed nouns, just like they would be when using Polish nouns. However, it must be emphasised that nouns in Group II are also allocated to masculine gender although they end in a vowel (or diphthong), whereas nouns in Group III end in a consonant, but are allocated to the feminine gender. If we take into account the proportions between Group I (65%) and Groups II and III taken together (35%), it becomes apparent that the children opt for the masculine gender based on the consonantal endings in 2/3 of cases, and in 1/3 of cases they are guided by a different factor. In sum, when it comes to allocating English nouns to Polish genders, the children do not seem to follow formal criteria in the way they might be expected to, despite the fact that many English nouns are phonetically close to existing Polish nouns (elephant-kant), and on that basis could be allocated to Polish gender classes. Other factors are also in the game.

\textsuperscript{55} IPA based
Group I (65%)  Group II (25.5%)  Group III (9.5%)

**Group I**

*C -> masc. (69)

1. taki snowman [snoomæn] 36. taki red flag [fleeg]
2. taki żywý snowman [snoomæn] 37. jaki taki red flag [fleeg]
3. biedny pen [pen] 38. na zielony flag [fleeg]
5. taki skeleton [skeletan] 40. frog pobrał z siebie [froeg]
6. lion szwedzki [laran] 41. ten dog [dog]
7. taki hen [hen] 42. mój bedroom [bedroom]
8. taki sun [san] 43. głodny worm [ws:m]
9. czerwony sun [san] 44. jaki break time [taim]
10. dolphin mój [dolfin] 45. ten dog [dog]
11. taki grass [gra:s] 46. moj bedroom [bedroom]
12. taki duży mess [mes] 47. na zielony flag [fleeg]
13. taki necklace [nekla:s] 48. mój sword [so:d]
14. ten juice [dzu:is] 49. my sword spadnieł [so:d]
15. taki cross face [feks] 50. caly custard brudny [kAstsid]
16. był ten fox [foks] 51. ten page [pædʒ]
17. od taki elephant [elefant] 52. taki carriage [kæridʒ]
18. takim flower pot [pot] 53. taki duży crunch [krAntJ]
19. ten picnic set [set] 54. ten witch [wiztʃ]
20. ten magic...tą magic [trumpet] 55. ten swing [swig]
21. ten magic trumpet [trumpet] 56. fajny swing [swig]
22. fajny hat [hat] 57. to duży swing [swig]
23. taki short bit [bɪt] 58. taki wheelbarrow [wi:əlbaˈrəʊ]
24. taki part [pa:t] 59. ten window [wɪndəʊ]
25. taki lift [lɪft] 60. ten rainbow [rɛntəʊ]
26. taki teapot [tiːpɒt] 61. niedobry apple [æpl]
27. ten teapot [tiːpɒt] 62. niedobry pineapple [pærnəˈpliːn]
28. duży clock [klokl] 63. pomarańczowy circle [sə:kəl]
29. taki book [bʊk] 64. drugi call [boːl]
30. taki book [bʊk] 65. taki ball [boːl]
31. duży duck [dʌk] 66. taki happy smile [smɑːl]
32. taki duck [dʌk] 67. taki move [mʌv]
33. taki beak [biːk] 68. taki hose [hoʊz]
34. ponzeżed speć snake [sneɪkl] 69. mój cup of tea [kʌp]
35. ten handbag [hændbæg]

**Group II**

*-C -> masc. (27)

1. ten scary spider [spɑːdə] 15. kolorowy tree [triː]
2. który jest tractor [trakta] 16. mały baby [ˈberbil]
3. duży big tractor [trakta] 17. ulepił mój mummy [ˈmʌmɪ]
4. jakis brown [kolor] [kɑːlə] 18. mój doggie [ˈdɒggi]
5. nudny jakié thunder [ɔnda] 19. duzego=duży dinosaur [dəˈzærən]
6. jaki number [nʌmbə] 20. ten baby dinosaur [dəˈzærən]
7. sticker tego [strɪkə] 21. taki dinosaur [dəˈzærən]
8. tuki magic mirror [mirə] 22. drugi go [goʊ]
9. ten monster [ˈmɒnstə] 23. mały boy [bɔx]
10. ten picture [ˈpɪktʃə] 24. taki śmieszny boy [bɔx]
11. atkilef mój willy [ˈwɪli] 25. fire, który [ˈfɑːr]
12. jak pony mój [ˈpeoʊni] 26. taki fire [ˈfɑːr]
13. mój monkey [ˈmɑːki] 27. taki zoo [zuː]
14. taki party [ˈpeəti]
In addition to codas, another candidate to guide the children in the choice of gender is animacy, regarded by linguists as "potentially one of the most profound conceptual distinctions available to children" (Gelman, 2001). The results of three different experiments conducted by Poulin-Dubois et al. (1996) suggest that children are able to discriminate animate from inanimate objects on the basis of motion cues by the age of 9 months. Although the children tested by Poulin-Dubois and her colleagues are monolingual, there is no reason to believe that this early-acquired ability would be very different for bilingual children, particularly those exposed to identical animacy rules governing gender systems they are acquiring. The question is whether there is any relation between the choice of gender and the animate/inanimate features in the mixed NPs. The natural gender marking doubled by the rules from the two languages render animacy likely to be a strong factor in gender assignment of the English insertions. Consequently, one expects feminine and masculine animate dimorphic nouns to be allocated to genders correlating with their natural gender. Masculine animate nouns in particular are expected to be marked consistently and correctly, since the dimorphism is supported by consonantal endings. The data shows that this is, indeed, the case, as all masculine animate dimorphic nouns included in mixed NPs are allocated to the masculine gender. One would also expect accurate correlation between the natural gender and the gender chosen for the feminine English insertions, since females can be explicitly feminine as a natural gender. The data shows, however, that feminine dimorphic nouns are assigned correctly in only 13% of cases. In all incorrect cases, they are allocated to masculine nouns. The examples below illustrate this situation:

Examples:

(77) Tam jest dużyduction
      there be.3P.PRES inside big-M duck
      'There is a big duck inside.'

56 nouns expressing the difference in form between individuals of different sex in the same species.
57 Animacy is a strong and an obvious factor, classifying nouns into natural genders. For that reason, the terms correctly/incorrectly have not been abandoned with reference to the animate English-Polish equivalents.
Both *duck* and *hen* may not be evident in their natural gender, but their Polish feminine equivalents are widely used for general reference. The masculine forms are used only when the male representatives are specifically referred to. It is puzzling, therefore, that the correct assignment of the feminine animate dimorphic nouns is only 13%. Dimorphism does not seem to be something the children draw on when allocating feminine nouns to gender, and due to such poor performance in the feminine, it is impossible to state with certainty if, despite the 100% correctness, the children are guided by semantics when dealing with masculine dimorphic nouns.

An additional variable to consider here is age. The data shows no age-related improvement in the accuracy for feminine dimorphic nouns, yet, interestingly, the accuracy for feminine non-dimorphic nouns, is far higher than in the case of dimorphic nouns. Moreover, a gentle increase in accuracy is found in feminine gender in non-dimorphic animate nouns. If we consider the fact that the semantics for non-dimorphic nouns is far less transparent than the one for dimorphic nouns, this finding is rather surprising. For masculine, the accuracy is most stable in the case of dimorphic and inanimate nouns. The percentage of accuracy is summarized in Table 25:

<table>
<thead>
<tr>
<th>Animate</th>
<th>Dimorphic</th>
<th>Masc</th>
<th>Fem</th>
<th>Neut</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100%</td>
<td>13%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Non-dimorphic</td>
<td>89%</td>
<td>42%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Inanimate</td>
<td>97%</td>
<td>16%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>

Following Corbett (1991: 10), “there is a massive amount of evidence...that controllers referring to animates are more likely to take semantically justified agreement than are those referring to inanimates”. This does not seem to be the case for the data in the mixed inventory discussed here. It must be remembered, however, that the mixed data discussed here is likely too small to allow absolute conclusions. We shall limit ourselves to saying that the gender choice and animacy seem to be two separate issues for the three children when dealing with the English insertions. In this particular pair of languages, where natural gender rules are present in both systems, we would expect animacy to over-power all other clues.
The fact that it does not support the children's choice of gender in the mixed data is somewhat disappointing. The results suggest that English nouns are treated as foreign to a larger extent than it had been expected, i.e. the gender assignment rules typically present when making gender choices for Polish do not seem to apply in the same way for English borrowings. In sum, when allocating English nouns to the Polish gender, the children do not show any preference for a specific gender form guided by animacy. At least, they are not guided by this feature in any clear way when faced with foreign animate nouns. Animacy plays little or no role in assigning genders to dimorphic English nouns to Polish genders.

By studying Czech speaking children Henzl (1975) found that they are much more likely to give preference to the least ambiguous forms over familiar forms. She showed the children a picture of a male person, at the same time labelling him as deda 'grandfather(M)', and observed that the children "produced a feminine inflection, analyzing the noun by its final sound as feminine, although the picture clearly indicated a male person" (1975: 194). An identical example of extralinguistic clues being overruled by less ambiguous intralinguistic morphological clues has been found in Jerzy's speech:

The child comments on a book containing pictures of female witch characters:

(80) Popatrz, ten witch is dancing.
    look this.M witch is dancing
    'Look, this witch is dancing.'

According to the data, Jerzy was insensitive to the natural gender both in Polish and in English not only around the time this example was produced, but also long before and after producing it. It is thus possible that ex. 80 is an amalgam of two factors: the child's disregard for the natural gender rule, and the consonantal ending of the noun witch. A word needs adding regarding translational equivalents. Let us consider the following example:

(81) Ulepi-l moj mummy.
    put.together-3SG.PST-M my.M mummy
    'My mummy put it together.'

_Mummy_ has been allocated to the masculine gender via agreement with not one, but two modifiers, i.e., the verb _ulepil_ 'put together/mould together' used in the past tense, and the possessive pronoun _moj_ 'my-M'. The semantic criteria are strongly overridden here. One could argue that the child may not have arrived at the stage when the sexual dimorphism in humans is obvious to her, but the database leaves little doubt as to whether Hania was aware of the gender distinctions found in the nature at the time of producing the gender mismatched borrowing. Animacy seems to be unavailable to the child at this moment, and it remains
unavailable for some time during this particular recording, as Hania produces a Polish noun phrase, in which she allocates the noun *mama* to the masculine gender:

(82) Co kopnq-l mama?

what kick-3SG.PST-M mummy(F)

‘What has mummy kicked?’

We could claim that these examples are isolated no attempt should be made at guessing what guided Hania in assigning masculine to such a blatantly feminine noun as *mama*. A possible explanation is that Hania treats the English insertion as a foreign word and thus allocates it to the masculine, which she retains for the Polish noun produced a few minutes after the mixed NPs.

Let us now turn to parts of speech in mixed NPs. Since the data is based on recordings from a very small group of children, and since these children are being brought up to minimize the mixing in their bilingual speech, it is difficult to establish the statistical significance of the regularities found in their database, which comprises of only a limited number of examples. It is of some interest to us that preliminary $\chi^2$ tests show no evidence that PoS marked for gender in mixed NPs play any role in the choice of the corresponding matching gender. However, if we work within the assumption that the children do not refer to the Polish equivalents anyway, the frequencies of PoS and the children’s preference for the attributive position may point to certain regularities, which may also refer to gender and gender assignment mechanisms.

English insertions permit alternative gender choices: one resulting from their formal structure, and one from their Polish equivalents. A problem of similar nature, i.e. of nouns such as *dziewçę* ‘girl(N)’ offering two gender choices$^{18}$, which can be established based the Agreement Hierarchy (Corbett, 1979), has been discussed in §1.0. According to the Agreement Hierarchy, the choice of agreement and also gender depends not only on the type of target (demonstrative, verb etc.), but also on its distance from the controlling noun: syntactic agreement is chosen with the PoS closest to the noun, whereas semantic agreement is chosen outside the NP. On that basis, it can be hypothesised that in mixed NPs, those PoS which are closest to the inserted noun will take syntactic agreement (the gender choice will be based on the structure of the noun), whereas those PoS which are further to the right from the controller are likely to take semantic agreement (gender will be more likely to be based on the semantics – i.e. translation equivalent). First, let us compare Corbett’s Agreement Hierarchy with the frequency scale for PoS in mixed NPs discussed in Analysis 2:

$^{18}$ It is semantically feminine, but morphologically neuter.
The collation of the two, Agreement Hierarchy and the frequency scale for PoS in mixed NPs, gives us an important finding. Since the attributive position means a higher likelihood of syntactic agreement, the children's preference for the modifiers occurring in this position shows their preference for the syntactic agreement, and is believed to be linked to their preference for the masculine gender. In other words, the frequency of the PoS might be a mirror reflection of what type of agreement the children choose most often. As it is the syntactic agreement, we can therefore say that they may be following formal clues, i.e. the consonantal endings. As a rule, consonantal endings guide Polish speakers towards the masculine gender. Monolingual data for the three children shows that they tend to overgeneralize that rule by treating feminine nouns ending in a consonant, such as *krew* 'blood(F)', as masculine. The majority of the English insertions end in a consonant, and the realisation of those endings can often be related to a Polish noun with a similar ending. It is likely that the children generalize noun endings in their speech first, and add semantic clues more productively later on. An example of such generalisation would be a noun whose meaning in Polish is well known to the children, yet when used as an English insertion the gender is chosen according to the phonological clues:

(83) To jest tak-i sun na niby.  
     it is such-M sun make-believe  
     'It's such a make-believe sun.'

Patrick's inventory gives evidence that he knew the Polish neuter noun *słońce* 'sun(N)' very well at the time of producing the utterance in ex. 83. The diminutive form *słoneczko* is also recorded in the boy's speech (2;7). The noun *słońce* is believed to have appeared in the input and the output between the ages 2;7-3;7, which gives reason to also believe that at the time of producing the mixed NPs *taki sun*, and allocating the English insertion to the masculine gender, Patrick knew the Polish neuter equivalent quite well, yet chose masculine.
The following section discusses a handful of English nouns that acquired inflectional affixes through contact with Polish. Due to the lacking evidence of agreement, they have not been included in any of the datasets discussed so far. The inflectional markers are, however, an indication of a placement in one of the Polish declensional classes, and it is worth examining how productive they are in the children’s speech. All three children, at some point, produced an inflectional suffix attached to an English insertion. The overall number is 31, and the individual proportions are: Hania 22, Patrick 6, Jerzy 3. Perhaps the most surprising observation here is that one of the two children who seemed to be treating English insertions “suspiciously” by opting for the safest, i.e. default, masculine gender, is generating those inflectional hybrids so productively. Less surprising is that the vast majority of the insertions produced by Hania received declensional suffixes that are characteristic of masculine declension class. It is nearly impossible to state anything regarding the production of the other two children, since their samples are minimal. Nevertheless, let us at least classify all these intraword switches and identify the declensional classes to which they may be allocated. First, the Polish declensional patterns will be presented to remind the reader the structure of the Polish inflectional system:

<table>
<thead>
<tr>
<th>Singular</th>
<th>II f.</th>
<th>III f.</th>
<th>IV n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>szkola (‘school’)</td>
<td>kośc (‘bone’)</td>
<td>wino (‘wine’)</td>
</tr>
<tr>
<td>Accusative</td>
<td>szkolę</td>
<td>kość</td>
<td>wino</td>
</tr>
<tr>
<td>Genitive</td>
<td>szkoly</td>
<td>koścį</td>
<td>wina</td>
</tr>
<tr>
<td>Dative</td>
<td>szkolę</td>
<td>koścį</td>
<td>winu</td>
</tr>
<tr>
<td>Instrumental</td>
<td>szkole</td>
<td>koścį</td>
<td>winem</td>
</tr>
<tr>
<td>Locative</td>
<td>szkole</td>
<td>koścį</td>
<td>winie</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plural</th>
<th>V masculine personal</th>
<th>VI non-masculine personal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>lekarze</td>
<td>koścį</td>
</tr>
<tr>
<td>Accusative</td>
<td>lekarzy</td>
<td>koścį</td>
</tr>
<tr>
<td>Genitive</td>
<td>lekarzy</td>
<td>koścį</td>
</tr>
<tr>
<td>Dative</td>
<td>lekarzami</td>
<td>koścį</td>
</tr>
<tr>
<td>Instrumental</td>
<td>lekarzach</td>
<td>koścį</td>
</tr>
<tr>
<td>Locative</td>
<td>lekarzach</td>
<td>kością</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 46. Polish inflectional system
None of the insertions in Table 26 gives us a straightforward indication regarding the gender class to which the nouns have been allocated (apart from the general indication, i.e. masculine). In other words, due to the overlapping suffixes, none of the masculine subgenders (animate/inanimate) can be identified. The suffixes pointing to the feminine gender are somewhat clearer: the accusative -q and the genitive -y are typical suffixes found only in the feminine declension. We could claim that feminine is identifiable from just those endings. However, if we consider examples such as *artysta*, which decline according to the feminine pattern, but are semantically masculine, such a claim becomes invalid. That is why evidence of agreement is required to identify the gender for those insertions.

Table 27 gives examples of English nouns, which acquired plural suffixes. In most of the examples, there is also evidence of agreement (PoS marked in italics). There are no examples of masculine personal gender. Example 84 below is the only example providing evidence of agreement with a noun that has acquired an inflectional suffix:

(84) Będzie wykluc *osię ma* mak-y pengwina-a.

`It will hatch (and it) has a small penguin.`
The adjective points to inanimate gender, but the inflectional -a in pingwina is typically an indication of animate masculine gender. The cross-marking that takes place in this example suggests that even with agreement of some sort, the English insertions may behave similarly to Polish nouns in Polish utterances, where children often struggle with the notion of animacy.

4.3 Conclusions

Learning the specific features of a gender system is a complicated task, since various aspect of the linguistic system are interlinked. The present study has shed some light on how bilingual Polish-English children deal with the requirement of the Polish gender system. It has shown that animacy and the availability of Polish equivalents play little or no role in the process of gender assignment of English insertions to Polish genders. Also, the data has shown that in the case of bilingual gender assignment to foreign nouns, it is more a question of finding a default gender (Hania and Patrick) or extending the rules used for the matrix language (Jerzy). Researchers have previously pointed out that “there can exist immense differences between the individual paths children take to arrive at roughly the same result, i.e. mastery of the full system” (Smoczyńska, 1985b: 616-617), and the three children investigated here have confirmed this claim. All three children show a bias towards masculine in the Polish and the mixed utterances, but their error patterns bear marks of individual development of gender. They eventually arrive at similar results, yet they exhibit fascinating differences along the way. The data investigated here have also suggested that in gender assignment of Polish and English nouns the three children may be using language specific mechanisms. In other words, in monolingual NPs they often use masculine, but feminine is active as well, whereas in mixed NP, masculine becomes a default gender, similarly to the default character of the English neuter for inanimate nouns.
Chapter Five. Results and implications

5.0 Introduction

Among the studies on the acquisition of gender there are very few that investigate acquisition of gender in the speech of bilingual first language learners. Only a few have studied children as young as Hania, Jerzy and Patrick, and many focus on school children, which does not allow reliable age-to-age comparison between them and this study. The studies that analyse the acquisition of gender by adults are not comparable at all, as the acquisition context is entirely different for these two groups. Thorough research of published and unpublished materials reveals that among studies that discuss bilingual acquisition of gender there are nearly none including Polish as one of the languages. This study fills this enormous gap in the child language research and offers an empirical contribution to research in at least four areas of linguistics: child language, morphology, language contact and bilingualism. It sheds light on the acquisition of grammatical gender by three pre-school children who are brought up in a balanced Polish-English environment, and provides an insight into not only child Polish, but also the morpho-syntactic interaction of Polish and English. The findings broaden the understanding of the ways the assignment system of a gender language works when confronted with linguistic material from a language not matching with respect to grammatical gender. Error patterns in the gender system have been analysed here, and tendencies in error production have been highlighted. Polish-English noun phrases provide unique material for observation of very young learners' ways of coping with bilingual material with respect to a category not shared by their two first languages. The aim of this final chapter is threefold. Firstly, it presents the findings of this thesis. Secondly, it discusses the impact of this study in the context of other studies. Thirdly, it lists ideas for future research based on the data gathered for this project.
5.1 Findings and implications

Let us first summarise the findings obtained from the analyses of gender errors in Polish. Gender in Polish is marked on many different elements and it is not difficult to find/elicit data on the children's gender-marking skills. To provide details of how the children are dealing with the learning of the grammatical gender, their gender errors have been classified and analysed. The error database has been divided into: errors in the singular, errors in the plural and errors across number. The most common are errors in the singular, and a similar tendency remains at least until the final stages of recording at the age of 4;1 (Hania) and 4;3 (Patrick). The most common replacement gender is masculine (both in monolingual data and in the mixed NPs), but feminine is also frequent. The two genders are used as replacing genders in two different ways. Masculine is chosen as the default very early on, and its status does not change over time. Feminine, on the other hand, is used more actively at early stages than at later stages. This conclusion is of utmost importance, because it agrees with findings from Czech, Russian and German. In the case of Czech, Lehecková (2000: 746) has shown that although correct masculine nouns are learned earlier, feminine agreement forms are used more productively in the first stage of the acquisition of agreement. Similarly, Popova (1973) showed that in Russian child language feminine gender is preferred in the early stages of the acquisition (Stage I predominance of feminine). German data provided by Mills (1986a) shows the overgeneralization of feminine articles *die* and *eine*. Also concerning German, more recent research conducted by Eisenbeiss (2003) and Hamburg Group has shown that German children initially have a two-way system: feminine vs. non-feminine, which is very similar to what the author found in her data for Polish-English children. When contrasted with neuter, feminine as a replacement gender was chosen more often than masculine, which suggests that masculine and neuter are initially regarded as one group (possibly due to declensional similarities). It is possible that although masculine is easier to acquire, in various languages the sequence of the agreement forms learned will vary depending on the salience and frequency of use. Feminine endings seem more salient that masculine endings and are therefore acquired earlier. It seems that irrespective of whether the acquisition context is monolingual or bilingual, the feminine vs. non-feminine system is non-language specific.

Error density indicates a “wave” tendency, in which the children were going through correct-incorrect waves every month. These error frequency waves show that the children’s attempts to come to terms with rules of the gender system may differ. Error frequencies show differences between the first and the second half of the recording period for Hania and Jerzy, but not for Patrick. It means that there is an underlying pattern for Hania and Jerzy, whereas for Patrick, the change in error frequencies is quite constant through the entire
recording period, indicating few changes in his approach to whatever errors he is making. Hania and Jerzy's error frequencies show that their attitude to gender situations in which they make errors changes with time in statistically significant way. The two children make choices, draw conclusions, and then change their choice (for better in Hania's case, and often for worse in Jerzy's case). Patrick seems to make choices, draw conclusions and remain in his conclusions for longer. The period of 18 months of recording sessions is certainly a time when they all battle with a number of aspects concerning their gender systems, and an their effort does not end with the final recording (3;10-4;3). Even Hania, who makes the fewest mistakes, at the age of 3;7 shows readiness to compromise her already grounded knowledge of natural gender, as well as past tense verbs form in 1st person singular.

Animacy rules in the monolingual data are noticed and begin to be implemented (at least partially) around the age 2;11-3;0, and before that time, animacy errors are very rare. There is little or no correlation between the frequency of animacy errors and the general frequency of animate and inanimate nouns in the output, which may indicate that the learning of animacy feature develops irrespective of the density of animate and inanimate nouns in the input. While such a correlation is lacking, another one is present. The two simultaneously increasing frequencies: of correct feminine NPs and incorrect animacy-based NPs indicate the children's growing readiness for and acceptance of animacy-based errors. Feminine agreement is becoming clearer, and at the same time, the children are feeling more capable of handling a new set of rules (animacy-based ones). This answers one of the questions posited in the final paragraph of §2.3: is animacy going to be compromised? If so, for how long? Animacy is compromised even at the age of 4;0, and certainly is compromised at earlier stages. For the three bilingual children studied here, both animacy rules and the semantic criteria underlying the gender system are of secondary importance. Their main efforts are focused on the masculine as the default, and only when they are ready, are more rules learned.

Errors in the plural appear after the age of 3;4. Unsurprisingly, non-masculine personal is more common as the replacing gender than masculine personal, which appears inconsistently at irregular intervals, and is not even partly assimilated at the point of closing the project. When masculine personal is used as a replacing gender, it is very often a personal pronoun oni, and not a modifier required in agreement situations. In most cases, non-masculine personal replaces masculine personal on past tense verb forms. The data show that just like masculine is the default in the singular, te 'these-N_M.PERS' is the default demonstrative in the plural. When referring to a plural noun with a pronoun, however, the children are guided by the pronoun frequency. As a result, in accusative plural NPs, the more frequent ich 'their-M.PERS' instead of je 'ich-N_M.PERS' is chosen. Other plural examples
suggest that the features related to number and case distinctions are acquired before the features of gender agreement. For instance, the children use a correct declensional suffix, but precede it with the wrong gender on the agreeing modifier; or a correct masculine personal nominative form constitutes a NP with a non-masculine personal indeterminate pronoun. Since the same process is repeated many times between the ages 2;8 and 3;10, it is suggested that the features related to case or number distinctions are learned (at least partly) before the features related to gender distinctions.

In general, there are few errors across number, although number errors as a separate group are made quite frequently. Masculine personal is nearly never replaced by singular genders; non-masculine personal modifiers tend to be replaced by masculine or feminine, whereas non-masculine personal past tense verb forms are often replaced by neuter. This finding indicates that the confusion between non-masculine personal and neuter, which stems from the paradigmatic similarity of the -e ending, may be long lasting, and may not be limited to that particular ending only.

One of the research hypotheses states that little evidence of cross-influence will be found between the Polish and the English gender system. This hypothesis has been confirmed. No dramatic increase in the gender-based errors has been observed at any time the English input equalled or exceeded the Polish input. This study has shown that animacy and the availability of Polish equivalents play a little or no role in the process of gender assignment of English insertions to Polish genders. This finding is somewhat surprising, especially in the light of the overwhelming research suggesting that if a subject is clearly male or female, the choice of gender will result from the animacy. This study also shows that in the case of bilingual gender assignment to foreign nouns, it is more a question of finding a default gender (Hania and Patrick) or extending the rules used for the matrix language (Jerzy). Both mixed NPs and Polish error NPs show similar trends: masculine is chosen most often. For Jerzy, this is simply an extension of a general bias that he exhibits in his monolingual utterances as well. In the mixed utterances, masculine is clearly over-represented in the speech of all three children. This might be considered an instance of doubt. When Hania and Patrick are in doubt about a Polish word (monolingual cases), their gender choice is largely based on the gender proportions they know in Polish (m: 53%, f: 37%, n: 10%). In their mixed utterances, the level of masculine gender is higher and cannot be explained from the proportions between the genders in Polish. Hence, for Hania and Patrick, doubt about the gender of a Polish word is not the same as doubt about the gender of English insertions. Jerzy is struggling with his gender assignment all the time. His strategy is to choose masculine as the default gender, both for Polish words (71%) and for English loanwords (88%). For Hania and Patrick, masculine is also a kind of default choice for
English loanwords (86%). The fact that Hania and Patrick are more established in gender in Polish (83% correct), but show bias towards masculine in mixed utterances may suggest that when they mix they choose masculine as the default to be clearer or consistent. Jerzy is biased towards masculine both in Polish and mixed utterances, which suggests that for him it does not make a difference whether the noun is Polish or English. Therefore, we have two different models emerging from those three children – Hania and Patrick (stable Polish, unstable mixed) and Jerzy (both unstable). Researchers have previously pointed out that “there can exist immense differences between the individual paths children take to arrive at roughly the same result, i.e. mastery of the full system” (Smoczyńska, 1985b: 616-617), which can be expected to be the case for children learning any language, not just Polish or English. One of the explanations proposed here for the dominance of the masculine is the fact that in mixed NPs, those parts of speech which are closest to the inserted noun will take syntactic agreement (a result gender will be chosen on the basis of the structure of the noun), whereas those parts of speech which are further to the right from the controller are likely to take semantic agreement (gender choice is therefore more likely to be based on the semantics - translation equivalent). An important finding here is that in the mixed utterances, masculine is the default for all three children, as opposed to only Jerzy in the case of Polish utterances. If the children learn gender by making judgements as to the agreement rules between the controlling noun and the target, English insertions could be gearing their choices towards masculine, thus intensifying the default character of masculine. This added influence, which is absent from the solely monolingual Polish context, is therefore bound to bias Polish-English children towards the masculine.

Let us now pay some attention to how this study can be interpreted in a wider context. According to the established rules of gender assignment (Corbett, 1991: 72), some borrowed words may be allocated to one of the regular declensional types and their gender then follows automatically. However, the phonological form of some English insertions found in this database does not fit into any of the Polish declensional types and so they remain indeclinable. The main assignment rules might thus be semantic and form-based, and the simplest hypothesis seems to be that English nouns are assigned to a gender the way borrowed nouns are, according to Corbett (1991). In her recent study, Chirsheva (2009) analysed gender assignment in Russian-English code-switches. The Russian-English speaking participants included in her project belong to different age groups (5-12 year-old school children, 20-22 year-old students, and 26-55 year-old adults), which does not allow for direct and detailed comparison, yet some general tendencies observed may be related to here. In Russian, indeclinable loanwords denoting non-human animates are masculine, whereas those denoting inanimates as neuter (Corbett, 1991: 72). Neuter is regarded as the
default for borrowed indeclinable nouns as their phonological and morphological shape differs from that of Russian nouns. Chirsheva (2009: 72) argues that if a noun is indeclinable, it seems to have no gender; it is neither masculine, not feminine, which means that it is neuter. The Polish-English data show there are a number of indeclinable nouns, which are frequent among English insertions, especially if they end in the codas that are not common for Polish declinable nouns (e.g., fire). The results from Polish-English and Russian-English data are similar, although not identical. Adults, students and children included in Chirsheva's experiment assigned the majority of English code-switched nouns to masculine gender, regardless of their form and animacy/inanimacy. Thus, Russian-English gender assignment of loanwords shows the dominance of masculine (60%), but also marks the presence of feminine (39%), and neuter (1%). The three pre-schoolers from the Polish-English context chose masculine as the default for both declinable and indeclinable borrowings. Contrary to Chirsheva's data, there is hardly any evidence of assignment to other gender apart from masculine, which has been assigned to 97% of indeclinable nouns. Only 3% of these nouns have been assigned to neuter, and non indeclinable nouns have been assigned to feminine, which could be explained with the mechanism described in the paragraph to follow – children chose feminine when they are absolutely sure it is the correct answer, at least on the vast majority of the cases. Indecinable nouns do not offer the phonological clue of the -a ending, and with this logic in mind, the children do not choose to allocate those nouns to feminine. All animate nouns in the Polish-English data have been allocated to masculine gender, regardless of the personal/non-personal status. Nearly all inanimate indeclinable nouns have been allocated to masculine, which shows that although their phonetic and morphological shape differs from that of Russian or Polish nouns, neuter is not regarded as the default for borrowed indeclinable nouns in both language pairs. Semantic analogy, as discussed by Chirsheva, is another way of assigning gender to lexical insertions (borrowings). The data used here, however, do not indicate any significant influence of the Polish equivalents when assigning English insertions to Polish genders. All Chirsheva's informants are much older than the pre-schoolers investigated here; thus the likelihood of conscious semantic analogy is much higher than in the speech of children between 2;5 and 4 studied here. Gender in Russian is shown by Chirsheva to be assigned according to a complex set of semantic, phonological and morphological rules, and two main strategies of gender assignment are: semantic and phonological analogy (Chirsheva, 2009: 85). Unfortunately, there is no in-depth information as to which word codas are referred to when the author concludes that phonological clues play an important role. This study has shown that the children also rely on phonological clues in gender assignment, but they show no interest in the neuter endings -o and -e when assigning English insertions to Polish.
genders. They have all overgeneralized consonantal endings as masculine, which may be interpreted as following the phonological rule of "consonant=masculine". As to the feminine ending -a, an interesting pattern observed in the data. When we look at the instances where especially Patrick and Hania are wrong for words ending in -a (although Jerzy seems erratic in his gender assignment, he does not spoil this pattern), there is a clear semantic reason: the words denoting an animal are clearly overrepresented here. Thus, inanimate nouns in -a and nouns denoting a female are generally realised as feminine. Animate nouns in -a are preferably realised as masculine. There seems to be a competition between two correlations in the children's data. The high correlation between animals and masculine (80% of words denoting an animal are masculine) encourages them to make all animals masculine, whereas the high correlation between -a and feminine encourages them to make words in -a feminine. Both correlations are available in the datasets, and as linguists, we know the correct outcome, but the children do not - choosing masculine for animals is a logical alternative. Especially as the children may have learned the close relationship between natural gender and grammatical gender for humans, sometimes even overriding phonological clues. In conclusion, the learning of the assignment to gender classes is believed to take place in two main stages: below the threshold, the children start with a default (usually masculine), and follow what they learned by rote learning (repetition to some extent does help here). Above the threshold, they are likely to overgeneralise the other gender (feminine), and in the meantime, they gradually learn the irregularities of the gender system.

Chirsheva's reason for studying the gender of the English code-switches is "to find out whether the same or different Russian gender assignment strategies are used to allot nouns from a certain non-gender language in the same language combination" (Chirsheva, 2009: 67). She explains that similar tendencies may mean that "there are certain features of gender assignment that are acquired conceptually, with variations depending on the degree of bilinguality and other socio-pragmatic or psycholinguistic factors" (Chirsheva, 2009: 67). Since her study includes a wider age group, such predictions are easier to make. The assumption that variations in the gender learning techniques may depend on the degree of bilinguality is a challenging and an interesting suggestion, which could be tested on different groups in different language settings. Adults with low bilingual proficiency might overgeneralise one of the genders due to the lack of knowledge of various declensional pattern. As a "playing-safe" strategy, they may also follow semantic analogy more often than more proficient bilinguals may. Children with various degrees in bilinguality would be expected to show less varied results due to their initially unstable gender distribution.

Research in developmental psychology, linguistics and cognitive science of the past 20 years suggests that pattern-finding process begins long before children discover such
subtle grammatical features as gender, which only suggests that the initial and later stages of learning a language may be happening within one unified technique (Single Process Model in the Usage-Based approaches), rather than within more than one technique (Dual Process Model in the UG approach). In other words, usage-based approach suggests no dichotomy between "core grammar" and "periphery" - all constructions may be acquired with the same basic set of acquisitional processes, namely "those falling under the general headings of intention-reading and pattern-finding" (Tomasello, 2003: 6). In such an approach, pattern recognition becomes the main perspective in the acquisition. Let us consider, however, whether there exists a clearly marked split between the two approaches as far as grammatical gender is concerned. Let us look at a child learning the English past tense marker -ed. Children's overapplications of -ed to irregular forms such as go (goed) are well documented in the literature. Generativists would say that children are equipped with innate knowledge of analyzing the morpheme -ed as past tense morpheme. They apply the ending productively, but even more importantly - also logically - by adding it to verbs, rather than nouns or adjectives (the application is not completely random). Therefore, children have some innate knowledge about what this particular ending is supposed to do in their utterance, i.e. mark a past tense activity. Errors result from having too limited amount of input. The same would be assumed about gender agreement: children are equipped with language faculty which analyses morpho-phonological properties of nouns, and by analogy attaches particular morphemes to adjectives, pronouns, verbs, etc. In Usage-Based, gender learning is a result of linguistic experience, during which children build abstract schemes that underline productivity. Since gender can only be recognised via agreement, the learning process is based on their recognising patterns such as "a noun ending with a vowel -a requires the same ending on neighbouring words". Correct feedback and communicative intentions strengthen their linguistic choices. However, could we not regard pattern-finding in Usage-Based as parallel to what is meant as by analogy in UG? Children's overgeneralisations are significant in both approaches. It seems that gender is acquired "by analogy" according to both generativists and cognitivists, which makes it difficult to make the split between them specific. On the other hand, two problematic areas can be pointed out with regard to the treatment of gender within UG. According to the UG approach, "core grammar" is the innate property of human mind, whereas "periphery" involves the lexicon (also irregular constructions, idioms, and pragmatics). The question arises: is gender a part of lexicon or the core grammar? If we regard gender as an integral part of every noun, then it is a part of lexicon, and consequently, according to the Dual Process Model, as a part of the periphery, it is acquired with the help of normal learning processes (similarly to one of the Usage-Based models, the connectionist model, which does not state any a priori rules). Yet, gender can
only be recognised via agreement, which is a syntactic process, and would therefore need to be classified as non-periphery. Does it place it in the core grammar? It is not entirely clear. Another question is: if UG children’s core grammar really is equipped with innate knowledge of analyzing morphemes, how do they cope with the fact that in fusional (inflecting) languages such as Polish, morphemes code more categories than one (which is the case for the above mentioned -ed ending in English)? Polish morphemes often code gender, case and number in one morpheme. Do children possess innate knowledge that guides them through all categories simultaneously to arrive at correct ending for agreement purposes? Again, it is also unclear. It seems, therefore, that models of learning gender predicted by the two approaches may be similar (analogy and pattern-finding), but by not assuming any a priori innate morpheme recognition system, usage-based approaches such as for example connectionism may be more straightforward.

Apart from pattern-finding as the core of acquiring the gender system in Usage-Based approaches, the results of this thesis also point to some additional methods that precede the moment when children begin to find patterns in gender system, or at least before there is tangible evidence of such a process. What is meant as the additional method is that fact that all the children have shown evidence of masculine being treated as the default gender\(^9\), and this choice is the first step towards the further acquisition of gender. It takes place within the pattern-finding process, because although it means that one of the genders is chosen “blindly”, such a strategy allows children to keep the majority or a number of nouns in the safe (default) group, and at the same time, they have a chance to observe other regular patterns present in the gender system. This usage-based approach is confirmed by the data collected and analysed here: all children choose masculine as the default, and meanwhile “work on” untangling the rules underlying the correct feminine agreement. They slowly discover that -a is a meaningful ending and that it requires agreement in -a with other parts of speech. They are initially quite careful about their “discovery”, and with time, more examples of correct feminine gender agreement are found, while masculine remains default for all the cases, which they are doubtful about.

Contrary to the unitary language system hypothesis, according to which early bilinguals go through a stage when they cannot differentiate their two languages, researchers have shown that children acquire language specific morpho-syntactic properties of the target languages, and these correspond to monolingual patterns (De Houwer, 1990, Döpke, 2000, Hulk, 1998, Meisel, 2001). No evidence has been found in this thesis of the participants’ mixing the English natural gender rule with the Polish rule of the three genders in the

\(^9\) The same tendency is found for Hanrican-Norwegian, Hanrican-Lithuanian, Hanrican-Portuguese (Weinreich, 1953) and Hanrican-Italian (Correa-Zoli, 1973).
It shows that Polish-English children also acquire language specific morpho-syntactic properties of Polish and English, thus confirming the previous research at least from the perspective of the acquisition of grammatical gender. This study also adds to the valuable collection of research on bilingual first language acquisition (BFLA), which includes research by such academics as Jürgen Meisel and his colleagues in Hamburg; Annick De Houwer in Belgium, Elizabeth Lanza in Norway, Marilyn Vihman in the US, and also work at McGill University in Canada. One of the initiators of the BFLA framework stresses the importance of bilingual studies in the following way: “Theories of language acquisition are currently based largely on monolingual children, but must ultimately incorporate the “facts” of BFLA if they are to be comprehensive” (Genesee, 2006: 45). BFLA gathers research on children who grow up learning two languages simultaneously, and most research has examined children who begin dual language acquisition at birth. It is believed that such acquisition provides the closest insight into the true bilingualism. This study provides data and observation to contribute its insights into this fascinating area of language studies.

### 5.2 Future research

The speech of the three children participating in this study constitutes a very rich source of data; hence, it would take many pages to exhaust all possible topics and perspectives available for analysis. As Genesee points out, “the study of bilingual acquisition is worthy in its own right” (2006: 45), it is believed that this work is not merely contributing valuable data, but also important insights and conclusions. Ideas and projects that may follow based on this database would contribute to the areas mentioned above, and probably many other fields, such as phonology or socio-linguistics. In the area of morphology, virtually any topic could be investigated: tense and aspect in the children’s developing verbs; the acquisition of case in Polish, and of number in both languages (the development of the concept of plurality); how children learn alternations so common in Polish morphology; forming of numerals; the learning of various prepositions and adverbs, etc. This thesis does not provide data on the consistency in the use of genders for the same noun, but investigating this aspect in the future would be a valuable extension of the work presented in this thesis. As far as the children’s lexical development is concerned, further work can be done on the rates of the early vocabulary development; also, word-formation and neologisms may be of interest. Understandably, all of these areas available for further analysis may be examined in comparison to monolingual Polish data available from other researchers. More topics to be investigated can be found in the area of socio-linguistics: how children and their parents
share the load of turn-taking in conversations; how topic-changing differs in bilingual and monolingual context, etc. The influence from English is not limited to the mixed utterances, as it also emerges in deviant syntactic structures found in Polish, making syntactic calques a promising area of study as well.

Due to the socio-economic changes of the last 20 years both in Poland and outside the country, there are more and more mixed language families with bilingually raised children, where Polish is one of the languages. The author is hopeful to plan and organise a wide-scale data collection from such families, so that in consequence, a large bilingual database of Polish acquired simultaneously with other languages can be created. Examples of language pairs that are already on the author’s list of possible candidates are: Polish-German, Polish-Italian, Polish-French, Polish-Japanese, Polish-Arabic. Such a scheme will require all types of support, which the author is positive will be available for such an important and up-to-data project.
References


