

**ANALOGY:**  
**THE RELATION BETWEEN LEXICON AND GRAMMAR**

**Iwona Kraska-Szlenk**  
**Warsaw University**

**April 2007**

## CONTENTS

Preface.....	iv
Chapter 1:	
<b>Introduction</b>	
1.1. The mechanism of analogy.....	1
1.2. Previous research on analogy as relevant to the present work.....	4
1.3. An outline of the present framework.....	7
1.4. The data sources, research method and the study's organization.....	16
Chapter 2:	
<b>The e~a/o alternation in Polish nouns</b>	
2.1. Preliminaries.....	20
2.2. The distribution of alternating and leveled nouns at different frequency ranges.....	23
2.3. Other factors enhancing/preventing analogy.....	28
2.4. An Optimality Theoretic analysis of the data.....	33
Chapter 3:	
<b>The o~u alternation in Polish nouns</b>	
3.1. The historical source of the alternation and its current scope.....	36
3.2. The distribution of alternating nouns within the lexicon.....	38
3.3. The problem of diminutives.....	42
3.4. Double diminutives.....	48
3.5. An Optimality Theoretic analysis of the o~u alternation.....	52
Chapter 4:	
<b>The e~∅ alternation in Polish nouns</b>	
4.1. The historical background and the current scope of the alternation.....	54
4.2. The e~∅ alternation in masculine nouns.....	58
4.2.1. The distribution of alternating masculine nouns within the lexicon.....	58
4.2.2. Analogy in nouns of the infrequent type.....	71
4.2.3. Analogy in masculine diminutives.....	75
4.3. The e~∅ alternation in feminine and neuter nouns.....	80
4.3.1. The distribution of alternating nouns within the lexicon.....	80
4.3.2. Multifaceted analogy in feminine and neuter nouns.....	84
4.3.3. The problem of the Base in feminine and neuter diminutives.....	92
Chapter 5:	
<b>Semantic distance and contrast: differences between nominal and verbal paradigms</b>	
5.1. Why should verbs and nouns differ?.....	101
5.2. Noun-verb asymmetries with respect to analogy in Polish.....	103
5.3. Inflectional patterns in Moroccan and other Arabic dialects.....	106

Chapter 6:

**Analogy vis-à-vis Zipf's frequency laws**

6.1. Zipf's laws – the background.....	114
6.2. Zipf's laws in Swahili morphophonology.....	115
6.3. Leveling in a verbal paradigm conditioned by Zipf's laws .....	125
6.3.1. The paradigm of the past tense of <i>ić</i> in standard Polish.....	125
6.3.2. The leveling of the [ʃ] stem in the colloquial language.....	128

Chapter 7:

**UR-driven analogy**

7.1. Introduction.....	131
7.2. Historical (in Yiddish) and sporadic (in Polish) loss of final devoicing.....	132
7.3. The development of Polish <i>deska</i> 'plank'.....	135

Chapter 8:

**Other issues**

8.1. The directionality of analogical mapping in derivational morphology.....	137
8.1.1. The problem of derived Bases.....	137
8.1.2. Regular versus analogical stress in English derivatives.....	138
8.2. Phonological universal constraints or language-specific morphophonological templates?.....	141
8.3. Concluding remarks.....	143

Annex:

The "final devoicing suppression" experiment.....	145
References.....	148

## PREFACE

This study examines the mechanism of analogy in the context of language use and from the perspective of the Optimality Theoretic formal model. I argue that language usage criteria, such as type and token frequency, underlie an abstract concept of “grammar”, but are not entirely synonymous with it. In the present proposal, the two are interrelated through a system of extended correspondence constraints, whose ranking with respect to each other and to markedness constraints represents “phonologization” of language use. The argument is supported by a detailed discussion of vocalic alternations in Polish using synchronic and diachronic evidence. The second part of this work concentrates on factors other than frequency which may cause or prevent analogical developments, but which are also motivated by language use. Illustrative linguistic material comes from a variety of languages including Polish, Swahili, Arabic and English. The study stresses the active role of lexicon in shaping language grammar. Due to the dynamic character of lexicon-grammar interaction, analogical changes are not only interpretable, but to some extent predictable from historical and synchronic data.

I would like to deeply thank all my colleagues from the Department of African Languages and Cultures at Warsaw University for their support when this work was being written. I am particularly indebted to Nina Pawlak for her encouragement and confidence in all my other projects and for discussing with me the earlier draft of this book. I am very much obliged to Piotr Bański for going far beyond proofreading and contributing valuable comments on various issues. I am grateful to my family for their patience and coping with me during the hard time of preparing this book and for making all efforts to be helpful and comforting. I also thank the persons who kindly agreed to participate in the experiment described in Annex and to my daughter Maja for her great enthusiasm to assist me in testing these and other Polish data.

## CHAPTER 1

# Introduction

### 1.1. The mechanism of analogy

The phenomenon of analogy has been known in the Western tradition since the time of the ancient Greeks, who recognized *analogía* ‘similarity’ as a mechanism operating in grammar. In the narrow sense of the term, “analogy” comprises two widespread processes known as paradigmatic leveling and four-part or proportional analogy. The former can be also called “stem analogy”, since it eliminates stem alternation within a paradigm. The Polish examples in (1) illustrate leveling in a declensional paradigm with respect to the *a~e* alternation, which in Early Polish was phonologically conditioned. In many lexical items, however, only one stem variant survived in the modern language. In the case of *czas* ‘time’, the *a*-vowel alternant replaced the original *e*-variant of the previous locative; in the case of *cena* ‘price’, the locative stem was mapped on nominative and other declensional cases. Sometimes leveling has a more general character and applies among related stem-sharing words. For example, Polish original \**sistrzeniec* ‘nephew’ became *siostrzeniec* under the influence of the word *siostra* ‘sister’, in which the *o*-stem allomorph occurs throughout the declensional paradigm (after having been previously leveled itself).

#### (1) Leveling in Polish paradigms

early paradigm	modern paradigm	gloss
ʦas (nom.): ʦeɕe (loc.)	ʦas (nom.): ʦaɕe (loc.)	‘time’
tsana (nom.): tseɲe (loc.)	tsena (nom.): tseɲe (loc.)	‘price’

Unlike leveling, which aims at stem uniformity, proportional analogy may introduce stem alternations, since it maps one morphophonemic pattern or “template” onto other lexical items. Sometimes the innovation affects words belonging to a certain inflectional type, reshuffling them into another class (e.g. Old English *cow*, pl. *kine* versus modern *cows*), hence some authors call it morphological change. The Polish example in (2) demonstrates how an early loan-word *wizerunek* ‘image’, which originally had an invariant stem form (attested in the 16<sup>th</sup> century, cf. chapter 2), changed its singular nominative according to the alternating pattern of other nouns of a similar phonological make-up, as shown by the example of the word *ranek* ‘morning’.

#### (2) Proportional analogy in a Polish paradigm

early paradigm	modern paradigm	gloss
vizerunk (nom.): vizerunku (loc.)	vizerunek (nom.): vizerunku (loc.)	‘image’
ranek (nom.): ranku (loc.)	ranek (nom.): ranku (loc.)	‘morning’

Even though the two processes sketched above may bring about opposite results with respect to the existence of stem alternation or lack of it, they are in essence very similar and each can be phrased in terms of the other. While leveling aims at identity of stem-related forms which share lexical meaning, proportional analogy aims at observed identity of forms sharing grammatical meaning. And conversely, while four-part analogy remodels a form according to a pattern found elsewhere in the language, leveling does the same according to a “stem non-alternation model” prevailing elsewhere in declension or conjugation. Let us observe in this context that leveling would not be possible if alternation characterized all lexical items of a given category, as it happens in languages with predominating templating morphology, as for example Classical Arabic.

The very nature of analogy, conveniently captured by the slogan “one meaning, one form” (attributed to Raimo Anttila), was recognized as early as in Wilhelm Humboldt’s work, as illustrated by the following quotation: “Since *words* always parallel *concepts*, it is natural for *related concepts* to be designated by *related sounds*. If the pedigree of concepts is more or less clearly perceived in the mind, a pedigree in the sounds must correspond to it, so that conceptual and sound affinities coincide” [Humboldt 1988:71, trans. from Humboldt 1836-1839]. To this philosopher’s credit, Theo Vennemann later coined the term “Humboldt’s Universal” and defined it as follows: “Suppletion is undesirable, uniformity of linguistic symbolization is desirable: Both roots and grammatical markers should be unique and constant” [Vennemann 1972:184]. Early authors underlined the psychological foundation of analogy and its “simplifying”, or perhaps better, “organizing” function, as for example expressed by Samuel Kroesch in the following statement: “Just as the basis of all phonetic laws is physiological, so the basis of all analogy is psychological. The association in the mind of one idea with another forms the basis of any analogical formation. Ideas are associated into groups in the mind, and so also are words which represent these ideas associated into groups. The tendency of analogy, then, is to counteract the great diversity in language and to bring the incongruous elements of speech into groups and systems, thereby simplifying them” [Kroesch 1926:35].

Looking from a different general perspective and trying to explain the directionality in analogy, a 19<sup>th</sup> century Polish scholar Mikołaj Kruszewski in a short, but remarkable article, connected the mechanism of analogical processes with that underlying phonological assimilation (Kruszewski 1879). In his view, assimilation (in the broad sense) involves elimination of a weaker element by a stronger one. Consequently, more salient or more frequent forms replace rarer, less salient ones in various analogical processes, including leveling, morphological change or folk etymology. M. Kruszewski died at a very early age in 1887 and had no chance to develop this thought any further, but it is possible that his research to some extent influenced his mentor and recognized linguist Jan N. Baudouin de Courtenay, as it is known that the exchange of ideas between the two went both ways. Commenting on analogy, Baudouin de Courtenay says: “More numerous forms are more likely to be preserved, forms, which are more recurrent in the language, forms constantly used, hence, those whose analogy predominates; for repeating of impressions [i.e. “expressions” – I.K.S.] makes them stronger and more durable. However, it is possible that a certain direction of analogy favors preservation of rarer

forms, and even creates new categories for them [1904/1974:399]” (translation from Polish I.K.S.). But the author does not discuss the problem any further.

As I will show later in this chapter, the modern views on the mechanism of analogy are not much different from the traditional ones exemplified above.

Neogrammarians juxtaposed analogy to sound change: while the latter was viewed as regular, affecting the whole lexicon in an unexceptional manner, the former was considered irregular and unpredictable in principle, operating in an item-by-item fashion. This clear-cut distinction between sound change and analogy was immediately and passionately questioned by Hugo Schuchardt, who pointed out numerous exceptions to the Neogrammarian unexceptionability of sound laws and presented a hypothesis of progressive sound change from a similar environment to another (cf. Schuchardt 1885/1972). Schuchardt and the slogan attributed to him that “each word has its own history” received some, though not many followers (cf. Malkiel 1967, Hock 1986, ch. 20 for discussion), and it was only much later when detailed linguistic evidence came from the research of William Wang and his associates in support of Schuchardt’s and *contra* the prevailing Neogrammarians’ views (cf. Wang 1977, among others). According to Wang’s theory of “lexical diffusion”, the effects of sound change are not simultaneous and lexically unlimited, but progress within time from one word to another, much in an analogy-like manner. With lexical diffusionists, the old Neogrammarian controversy revived. To complicate the picture, William Labov’s research, started in the 1960s and conducted for many years since then, proved the extreme importance of social factors in the propagation of sound change (cf. Labov 1972, 1994, among others). Labov himself proposed a conciliatory solution to the issue of the mechanism of sound change, demonstrating that its character, along the regular or lexical diffusion lines, depends on the kind of change in question and relates to such properties as discreteness, abstractness, as well as grammatical and social conditioning, with changes of more “phonetic” type oscillating at the “regularity” extreme and more abstract, “phonological” changes being more “diffusionistic” (cf. Labov 1994). Hock (2005) points out that Labov’s continuum may be treated as a part of a larger continuum of changes of basically analogical character whose different behavior relates to a potential domain of applicability: the broader the domain, the greater the regularity. Hence, sporadic changes as e.g. recomposition, blending or contamination have very narrow domains of potential applicability, while changes involving minimal phonological restrictions, such as, for example, British English *r*-insertion, are relatively regular. And four-part analogy and leveling are fairly systematic, since they typically involve inflectional classes rather than single lexical items.

As the identification of “analogical change” may vary considerably depending on one’s world view, similarly a model of synchronic grammar may recognize importance of analogy to different degrees. For example, in the early generative framework, synchronic analogy was not acknowledged at all and even as a historical process it had a very low status, crammed into the overall system of rule addition, simplification, reordering etc. (cf. Kiparsky 1968, 1978, King 1969, as well as Anttila 1977, ch. 4 for a critical review), even though a need for surface-oriented “paradigm uniformity” was observed (e.g. Kiparsky 1972). A more recent generative model of Optimality Theory (OT) distinguishes analogy as a separate phenomenon by means of ‘identity’ constraints particular to it, as discussed later in this chapter. In cognitive linguistics, analogy in the

broad sense of similarity and pattern extension underlies the essential ideas of the framework, such as “conceptual metaphor” (Lakoff and Johnson 1980) and “schematic structures” (Langacker 1987, 1991). It is also possible to construct a model of grammar with analogy as the only formal device, as for example in Skousen (1989).

From the very broad perspective of reasoning and learning in general, analogy, as Raimo Anttila puts it, “mediates between actuality and potentiality” [Anttila 2005:426] and thus has enormous power in various domains of human activity. As the author continues: “Humans are simply analogical animals. Language structure and language use are also predominantly analogical and this is why analogy is the backbone of universal grammar” [Anttila 2005:438n.]. This statement may seem exaggerated and too strong. But only at first. Not if we take a moment and realize that “language structure” is far from a nicely written reference grammar and not if we take language use seriously. If we do, we can see quite clearly that “language” is a great chaos and that analogy somehow makes its way through this chaos, slowly, effectively and all the time. Much of the present work will be devoted to showing just that.

In the following section, I will return to the issue of “narrow” analogy and its analysis in more recent times. Reviewing all the vast literature on the subject is of course impossible and I will make my own subjective choices. Anttila and Brewer’s (1977) bibliography can be consulted for earlier works and a fuller account of modern references, especially in the Optimality Theoretic framework, can be found in Albright (2002, 2005) and McCarthy (2005).

## **1.2. Previous research on analogy as relevant to the present work**

The Polish linguist Jerzy Kuryłowicz was the first to formulate systematic universal “laws” of analogical change, cf. Kuryłowicz (1947, 1964). His proposal was soon followed by a polemical response from another Polish linguist, Witold Mańczak, who recognized universal “tendencies” (later renamed as “laws”, too), cf. Mańczak (1958, 1978, 1980, 1996). An excellent detailed review of the “Kuryłowicz-Mańczak controversy”, as it is sometimes referred to in the literature, can be found in Hock (1986, ch. 10). My own discussion of it will be limited only to the points relevant to the present work. Kuryłowicz’s laws of analogy are briefly stated in (1).

### (1) Kuryłowicz’s laws of analogy

- I. A bipartite marker tends to replace an isofunctional simple marker.
- II. The directionality of analogy is from a “basic” form to a “subordinate” form with respect to their spheres of usage.
- III. A structure consisting of a basic and a subordinate member serves as a foundation for a basic member which is isofunctional but isolated.
- IV. When the old (non-analogical) form and the new (analogical) form are both in use, the former remains in secondary function and the latter takes the basic function.
- V. A more marginal distinction is eliminated for the benefit of a more significant distinction.

- VI. A base in analogy may belong to a prestige dialect affecting the form of a dialect imitating it.

Among Kuryłowicz's six laws, five have rather small domains of application and only the second one has a very general character. In fact, it is perhaps too general and a little imprecise because of the very open notion "sphere of usage". It may comprise cases of morphologically complex/non-derived words, rare/frequent forms in terms of text occurrence, or marked/unmarked elements, as for example, nominative is "unmarked" with respect to other declensional cases, or 3<sup>rd</sup> person singular is unmarked in an inflectional paradigm (cf. also Greenberg 1966). Kuryłowicz's second law will be important for my own analysis, although, as I will show momentarily, all different understandings of "sphere of usage" are better captured and unified under Mańczak's approach. As to the remaining laws, the fourth one will be seen at work in a few cases discussed in chapter 2, but I will argue that it follows from more general principles<sup>1</sup>. Some other cases treated in chapters 2-4, in which leveling affects "unmarked" nominative, could be seen as an instance of the third law, but they will receive a different interpretation under my account. Other laws will be immaterial for the present work, since no relevant data will be discussed.

While Kuryłowicz was more concerned with morphology and proportional analogy, Mańczak puts stress on phonological developments and leveling, discussed in the context of language use. Eleven detailed "tendencies", comprised in the author's later work into four general "laws"<sup>2</sup> (e.g. Mańczak 1980, 1996, ch. 7), are presented in (2). The first law appears as a "repair strategy" to eliminate alternations introduced by phonological development and corresponds to Humboldt's Universal. The second law counteracts reductions often caused by regular or irregular phonological development. It involves the same idea of greater salience of a linguistic element as Kuryłowicz's first law, but is more general. Laws III-IV are strongly tied up with Zipf's (1935) statistical laws relating frequency of a linguistic unit and its size or complexity, resulting in more frequent forms being better memorizable and more stable. These two laws parallel Kuryłowicz's second law, except that the base of analogy is clearly indicated as more frequent (hence shorter), i.e. more salient. Laws II and III may seem to contradict each other, but in fact they refer to distinct cases, as for example replacing a suffix with another (longer one) – law II, and leveling a (shorter, typically more frequent) stem variant in a paradigm – law III.

(2) Mańczak's laws of analogy

- I. The number of morphemes having the same meaning more often diminishes than increases.
- II. As to shorter/longer morphemes, shorter/longer words, words/word groups, the latter more often replace the former than the reverse.

---

<sup>1</sup> NB exceptions to this law are quite common, cf. Kiparsky (1974) or Hock (1986, ch. 10).

<sup>2</sup> Mańczak (1980:285) mentions a fifth law, too, which is not included into general analogy laws in the more recent publication of Mańczak (1996), presumably because it constitutes a specific case of the fourth law. It says: As to the locative case of geographical names/common names, non-locative cases of common nouns/personal names, the former keep an archaic character more often than the latter.

- III. As to shorter/longer morphemes, shorter/longer words, words/word groups, the former remain more often than the latter, keep more archaic character than the latter, cause changes of the latter than the reverse.
- IV. As to more frequent/less frequent forms, the former remain more often than the latter, keep more archaic character than the latter, cause changes of or replace the latter than the reverse.

In addition to the above laws of analogy, Mańczak formulated another general law as an explanation of a whole range of phenomena. It says that linguistic units which are used more frequently are typically more differentiated than rarely used linguistic units (cf. Mańczak 1966, 1996 ch. 8 and 9). The law accounts for cases of suppletion, as well as various instances of unmarkedness in the sense of Greenberg (1966). In its application to analogy, it elucidates a tendency to reduce pattern allomorphy in rarely used or small categories. This “differentiation law”, as well as the fourth of Mańczak’s analogy laws will be most relevant to and supported by the present study. His first and third laws will be shown to follow as a consequence of them rather than being independent principles. As to the second law, it has no application to the data treated in this work.

The idea connecting analogy to frequency of use is scattered throughout various other works on analogy (including general publications as e.g. Anttila 1989 or Hock 1986), but it has the most systematic expression in the research of Joan Bybee (e.g. Hooper [Bybee] 1976, Bybee 1985, 1998, 2001, Bybee *et al.* 1994). In Bybee’s dynamic model, the lexicon and grammar “emerge” out of recurring patterns of language usage. The higher the frequency of a given lexeme, grammatical morpheme, fixed phrase etc., the greater its salience, leading to a stronger mental representation (cf. also Langacker’s 1987, 1991 “entrenchment”). All stem-sharing or affix-sharing units are associated with each other into a network, establishing lexical, semantic and grammatical connections, with weaker items linked to stronger ones. Bybee’s usage-based approach models the directionality of analogy in the simplest possible way by establishing a new connection from a poorly represented item or pattern to a stronger one. It is important to note that “strength” may result either from high token frequency, or from high type frequency. Like Mańczak’s, Bybee’s ideas concerning the analogy-frequency connection underlie my own approach outlined in the following section.

Optimality Theoretic framework (Prince and Smolenski 1993/2004) made it possible to introduce analogy into the architecture of generative grammar by means of “identity” constraints, formally expressed as a correspondence relation between various output forms. Starting with the earliest works, such as Benua (1995, 1997), Kenstowicz (1996), Kraska-Szlenk (1995/2003), McCarthy and Prince (1995), in which Humboldt’s Universal was reformulated in OT terms, most of the research of the past decade has concentrated on the issue of the Base in the correspondence relation. While there has been unanimous agreement as to the fact that morphologically derived forms have less complex words as their Bases, there has been no consensus in the domain of inflectional paradigms, whose members are not morphologically derived from one another. The problem involves fundamental questions such as: Is there a unique form in the paradigm which serves as a Base, or are multiple Bases possible as well? Or, perhaps there is no Base at all and the directionality of leveling in a paradigm follows from something else?

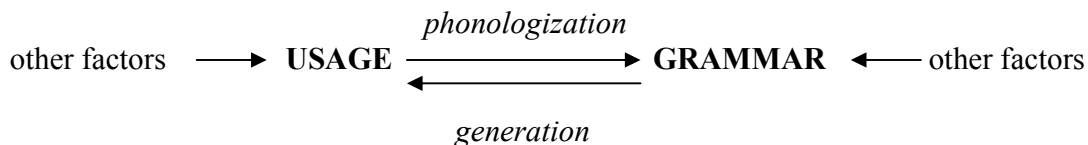
Most authors agree that a highly ranked phonological constraint may enforce the directionality of analogy and promote an allomorph which better satisfies markedness. But in some cases of analogy, none of the two (or more) allomorphs complies better with the phonology of a given language and such strategy may not apply. Especially problematic are “split” situations, in which one stem allomorph is leveled in some lexical items and another one in another lexical (grammatical) class. To deal with such cases some authors propose solutions reminiscent of Kuryłowicz’s second law and appeal to unmarkedness of certain forms, as e.g. Kraska-Szlenk’s (1995/2003) analysis of Polish diminutives leveled to the nominative case, or Kenstowicz’s (1996) example of Spanish verbs leveled to the indicative mood. A different line of reasoning is proposed by Albright (2002, 2005), according to whom leveling takes place towards a form which is maximally informative as to the underlying representation (cf. chapter 7 for a more detailed discussion). McCarthy (2005) rejects the notion of the Base in inflectional paradigms altogether, opting for multiple correspondence among all forms and the majority criteria decision making (cf. chapter 5 for discussion). Multiple correspondence among various stem-sharing forms is also proposed by Steriade (2000), who claims that analogy may involve mapping of various parts of structure from different Bases. Contrary to these proposals, Albright (2002, 2005) argues for a strictly one-Base analysis.

Although the OT views on analogy briefly sketched above may seem quite different and incompatible with each other, I will draw to some extent on all of them, as discussed in the next section.

### 1.3. An outline of the present framework

My own analysis of analogy will be based on a systematic description of language use data with all their details and nuances. I assume that language usage motivates and underlies a language grammar, but should not be automatically equated with it, or automatically inferred from it. As Frederick Newmeyer candidly says in the title of his article (Newmeyer 2003): “Grammar is grammar and usage is usage”. But the connection between the two is too close to be ignored. I assume that language usage and grammar mutually influence one another, as represented by the arrows in the sketchy diagram in (3). Phonological rules (constraints, etc.) of grammar are themselves derived from usage, but they also generate it. In this way, each component constitutes an input to, but also an output of the other. Since other factors (e.g. social, cultural, historical, political, geographical, etc.) have an impact on usage, as well as on grammar independently, the language is in constant fluctuation, with its usage and grammar parts being always a little mismatched.

#### (3) Usage and grammar in language



Some changes are usage-based and some are grammar-based. In the case of analogy, the vast majority of changes seem to belong to the former kind and few to the latter. Among usage-induced changes, those that are triggered by frequency are particularly common. Therefore, frequency appears as the most important factor in analogy, although other factors may occasionally take over. To use a comparison, frequency-driven analogy seems as natural and widespread as, for example, assimilation or lenition in phonology, while analogy due to other causes is as irregular and uncommon as, for example, dissimilation or metathesis. But this does not mean that the latter does not exist. And thus the present work, which examines the way analogy operates in a portion of Polish lexicon (chapters 2-4), will demonstrate vast evidence for analogy determined by frequency criteria and only few “exceptions” to it. It is only through a detailed examination of all cases in a given domain that we can see such a result and never through selecting the data of a particular kind from various domains.

We can talk about “phonologization” of usage, when a given change determined by frequency criteria is regularized as a part of grammar. In the OT framework adopted in this work, an analogical change will be modeled as an upward/downward movement of the given output-output (O:O) correspondence constraint (Cor). Its upward promotion with respect to some phonological (or morphophonological) constraint(s), previously responsible for an intraparadigmatic alternation (Alt), pictures leveling, as schematized in (4a). Its demotion below the “alternation” constraint(s) characterizes pattern (proportional) analogy, as in (4b). (The symbol “>>” marks the relation of constraint dominance, and the arrow indicates the direction of the change.) From the synchronic perspective, these two kinds of analogy are best visible, when a certain lexical (or morphological) class A of the given category follows stem analogy (no alternation present) and another class B of the same category follows pattern analogy (surface alternation)<sup>3</sup>, as schematized in (4c). The exact formulation of correspondence constraints will be discussed later in this section.

(4a) Constraint reranking in leveling

$$\text{Alt} \gg \text{Cor} \{O_x:O_y\} \quad \rightarrow \quad \text{Cor} \{O_x:O_y\} \gg \text{Alt}$$

(4b) Constraint reranking in pattern analogy

$$\text{Cor} \{O_x:O_y\} \gg \text{Alt} \quad \rightarrow \quad \text{Alt} \gg \text{Cor} \{O_x:O_y\}$$

(4c) Synchronic constraint ranking in stem and pattern analogy

$$\text{Cor-A} \{O_x:O_y\} \gg \text{Alt} \gg \text{Cor-B} \{O_w:O_z\}$$

A synchronic situation schematized in (4c) may result from a diachronic process as either (4a) or (4b), and is dynamically evolving in the same direction. This means that if leveling already took place with respect to some lexical items, we expect that other lexical items may also be affected by it, i.e. reshuffled from class B to A in (4c). And conversely, if pattern analogy introduced paradigmatic alternation into non-alternating

---

<sup>3</sup> I should add that this latter case would not be treated as “analogy” in a typical generative analysis.

words in the past, we can expect that other lexical items may move from class A to B. In the following chapters of this work, I will demonstrate that this kind of analogy dynamics is strongly correlated with a number of factors acting simultaneously, such as:

- sizes of classes A and B within a given category
- text frequency of particular members of classes A and B
- semantic distance between the corresponding members within each class.

The first of the above criteria, also known as type frequency, is understood strictly in the dictionary sense, i.e. reflects the number of words in the lexicon that share certain morphophonological features. In the sketchy example of (4c), this amounts to how many members belong to class A or B. In general, the more members in a class, the more salient it is. If such a salient class happens to be alternating, as B in (4c), it may attract other members, especially if class A is small. Nevertheless, it is important to consider pattern frequency in relation to text frequency of the members of the class, in particular, whether it is represented in high frequency ranges. It may happen that even a large class will disfavor stem alternation and undergo leveling, if its members are found only among low frequencies, since it would be equally difficult to learn the alternation by rote as to derive it by rule without having memorized words of high frequency which provide a pattern to follow.


The second criterion, also referred to as token frequency, indicates how often a particular word occurs in running text. High frequency words tolerate alternation better, because it is easily memorized through constant repetition. Therefore, in the previous example of (4c), members of class B will generally occur in text more often than those of class A. However, the expression “token frequency” is ambiguous in the case of an inflectional language, since it may refer to a lexeme as a whole, i.e. to occurrences of all possible forms of the given word, or its occurrences in one particular form. In this work, I will take into account both of these interpretations, distinguishing them by the convention of using the term “WORD” (in capitals) to refer to the former, and the term “word-form” in reference to the latter. For example, the frequency of the Polish noun LAS ‘forest’ indicates the sum of frequencies of all word-forms of the paradigm, such as *las* (nom. sg.=nom. acc.), *lasu* (gen. sg.), *lasy* (nom. pl.), etc. WORD frequency situates the given lexeme as belonging to common or rare vocabulary and its value determines the position of this lexeme in frequency dictionaries. Nonetheless, for the purpose of examining analogy, frequencies of particular word-forms are usually more informative, especially when a given stem alternant is underrepresented in the paradigm. Suppose that it occurs only in one declensional case, for example the locative, as in certain paradigms of Polish nouns discussed later in chapter 2. The relative frequency of such a word-form with respect to other cases may vary greatly depending on the meaning of the word. For example, names of places have unusually high frequency of the locative case. Relative frequency of a given stem allomorph within a paradigm becomes particularly important for rare words, because it triggers the directionality of analogy.

Each small change affects quantitative relations within the lexicon, which in turn may trigger other similar changes. If leveling takes place in one word of a given alternating class, it diminishes the size and “strength” of this particular class, which makes other lexical items more susceptible to leveling, too. And vice versa, each case of

pattern analogy makes the alternation more salient, which may enhance its diffusionistic spread on other words.

The third of the above criteria highlights the role of meaning in analogy, portraying Humboldt's Universal in a more fine-grained fashion. In a language with rich morphology, stem-sharing words often constitute large "families", in which all words semantically connect to one another to various degrees. The semantic correspondence has a gradient, non-discrete character, as illustrated in (5) by the Polish noun *kwiat* 'flower-nom. sg.' and some other nouns with this stem, arranged on a scale from the closest to the most distant semantic connection. Although this ranking reflects my own native speaker intuition and may seem somewhat arbitrary in the middle, there should be no doubt at least to its edges, with the smallest semantic distance found among word-forms of the same declensional paradigm, and the greatest distance between 'flower' and highly lexicalized 'April'.

(5) The semantic distance between *kwiat* 'flower' and morphologically related nouns

kwiat : kwiatu 'flower-gen. sg.'	smallest distance
kwiat : kwiaty 'flowers-nom. pl.'	
kwiat : kwiatek 'flower-dim. nom. sg.'	
kwiat : kwiaciarka '(lady) florist-nom. sg.'	
kwiat : kwiaciarnia 'florist's shop-nom. sg.'	
kwiat : kwietnik 'flower-bed-nom. sg.'	
kwiat : kwiecień 'April-nom. sg.'	

The greater the semantic distance, the smaller the pressure for analogy and vice versa. Hence, leveling within inflectional paradigms is most common, and semantically transparent derivatives are more likely to undergo analogy than more distant ones. For example, out of the derivatives in (5), *kwiaciarka* '(lady) florist' and *kwiaciarnia* 'florist's shop' have analogized forms which copy the vowel *a* of the word *kwiat* 'flower', unlike two other more distant derivatives with the stem vowel *e*.<sup>4</sup> This kind of analogical development does not relate to usage, but takes place directly in the grammar, due to the speakers' need to tighten the semantic connection, so that the meaning of the derived words will be transparently decomposable as 'flower+agent (fem.)' and 'flower+place (shop)'. Semantic distance also explains many cases of Kuryłowicz's fourth law, cf. the Polish new locative *czole* 'forhead-loc.' analogized according to the other members of the declensional paradigm, and the old form preserved in the semantically distant expression *na czele* 'in the forefront (of e.g. a parade)'.

Let us observe in this context that while frequency and semantic distance are both gradient in usage, their phonologization through appropriate ranking of correspondence constraints has a discrete character. Unlike phonetic changes, which typically proceed gradually, with intermediate stages of e.g. vowel quality, analogy never produces in-between forms which would reflect their real frequency or the degree of their semantic connection. The old *\*kwieciarnia* changes into *kwiaciarnia* in one step, although at the time the change is taking place, the two forms occur in variation.

<sup>4</sup> The transparent meaning of the two analogized words is due to the clear semantics of their derivational suffixes, unlike in the case of the next two words in (5).

Frequency-driven phonologization does not proceed automatically and needs real time to implement. Therefore, there will always be peripheral forms, still belonging to an old alternating category, even though by frequency criteria they should be already in a new, leveled one. Within the stem-sharing family illustrated earlier in (5), there is one clear example of such an exception, viz. *kwiecie* ‘flower-loc.=voc.’, whose frequency is low enough to undergo analogy (i.e. become *\*kwiacie*), but which is still in use in the old form (cf. chapter 2 for discussion). Exceptions of this kind are rare and are expected to be such, because even though grammar is one step behind language use, it follows in its footsteps. On the other hand, we also expect some resistance to automatic operation of analogy, because language belongs to a social and cultural domain. Various factors, such as the prescriptive influence of the older generations on the speech of children taught to speak “correctly”, school, literature, theater, etc. may help to maintain an unproductive, recessive alternation. We also come across sporadic changes which for prestigious reasons are directed *against* frequency. For example, Długosz-Kurczabowa and Dubisz (1999:57) point out that the frequency of the nominal suffix *-ość* in Old Polish was not as high as expected, due to its occasional replacement with *-stwo*, perceived as more elegant. According to Hentschel (1996), an “anti-Mañczak law” replaces semiotically “used-up” forms and operates under the conditions of linguistic non-conformism being in opposition to the usual “conformism”. One of Hentschel’s examples includes the analogical spread of the nom. pl. *-a* ending on Russian masculine nouns, e.g. *gorodá* ‘towns’, *glazá* ‘eyes’, *professorá* ‘professors’, clearly not triggered by frequency and, moreover, affecting mostly high frequency nouns, so that the innovation would be more conspicuous. Finally, frequency of a particular stem allomorph is irrelevant in the case of the so-called “UR-driven” changes discussed in chapter 7.

The existence of cases of the kind mentioned above in no way undermines the role of frequency in analogy – it only shows that language grammar is not ideally determined by it, as was already pointed out earlier. To sum up, an adequate model of grammar should be abstracted from usage to a large extent, but at the same time it should be flexible enough to accommodate the data which go against usage dynamics. It should be also sufficiently spacious to account for the finest distinctions among various related words, as exemplified earlier in (5). The model which I will propose stems out of Optimality Theoretic architecture, but it expands and modifies the standard version of correspondence (McCarthy and Prince 1995), and can be therefore called “extended correspondence model”.

Similarly to Joan Bybee’s framework of lexical connections, I assume that all stem-sharing words are in correspondence relations with one another as to their meaning and form. I further assume that each word potentially serves as a Base<sup>5</sup> in a phonological correspondence constraint, which has therefore a unidirectional character. The notation: Cor- $\{X:Y\}$  reads: “X must correspond to Y” and not vice versa. Since each existing word constitutes a possible Base, there is a potential constraint Cor- $\{Y:X\}$ , as well. Assuming that X and Y are two words predicted by Alt to have different allomorphs of the same stem, and provided that no higher markedness constraint intervenes, the ranking between respective Cor constraints will decide which allomorph will become the actual Base, i.e. which structure will be leveled under analogy, as schematized in (6) below. Naturally, the

---

<sup>5</sup> I follow the OT tradition in spelling “Base” in the correspondence relation with a capital letter (and I use “base” when talking about morphological base, so that the two are not confused).

mirror-image dominated Cor constraint is invisible – its work is vacuous. Since the exact interpretation and evaluation of unidirectional correspondence presents a rather complex issue, I will postpone the discussion of it until later in this section. For the moment, let us tentatively assume that the intuition of the Base having a phonologically predicted structure is secured one way or another. It should be added that if X and Y are real inflected words, there should be differences among them resulting from having different grammatical affixes. To guarantee that a correspondence constraint is satisfied to the extent that a phonological shape of the common morphological structure is mapped only and not the whole word (i.e. with an affix), I assume that the identity requirement is ascribed to the shared stem (root) of the words (as well as larger units, such as compounds, fixed phrases etc.) being in a correspondence relation (as e.g. in Kraska-Szlenk 1995/2003 or McCarthy 2005), and not to the entire words, as in some other OT analyses. (The latter is possible to maintain on the assumption that highly ranked morphological constraints of the type: “such and such case/person/tense etc. must have such and such affix” dominate Cor, so that the affixal part of the word’s structure can remain unaffected by it.)

(6) Cor constraints’ decision making

Cor- $\{X:Y\}$ $\gg$ Alt, Cor- $\{Y:X\}$	Y’s-allomorph copied
Cor- $\{Y:X\}$ $\gg$ Alt, Cor- $\{X:Y\}$	X’s-allomorph copied
Alt $\gg$ Cor- $\{X:Y\}$ , Cor- $\{Y:X\}$	X and Y (no leveling)

The representation of grammar in the form of extended correspondence constitutes an extremely powerful mechanism and therefore must be reasonably restricted in order to be something more than a notational device. I will return to this issue later in this section. But before that, let us observe that output-output correspondence highly reduces the need for input-output constraints (I:O) and approaches a one-level model of an OT grammar (cf. e.g. Burzio 2005, Myers 1999, Russell 1995). Thus, the extra cost of extended, omnipresent O:O correspondence can be compensated by the absence of the respective I:O constraints. Similarly to Burzio (2005), I assume that the OT constraint hierarchy essentially “checks” the surface forms. However, unlike Burzio and some other authors, I do not completely reject the notion of underlying representations (URs), but rather assume a “soft” version of them. For one thing, this means that the “hard” version, i.e. abstract, maximally underspecified URs are often quite unnecessary, unrealistic or even flawed (see e.g. Bybee 2001:20-21 or Burzio 2005 for argumentation). Hence, the surface output equal to the underlying input should be preferable whenever possible, the principle known in OT as Lexicon Optimization (Prince and Smolenski 1993/2004). On the other hand, the concept of the UR as a mental image unifying different allomorphs sometimes seems very appealing and more realistic than its absence. What I mean in particular are cases of morphemes traditionally represented in their URs as floating features, which are fused on the surface with various other morphemes (cf. e.g. Akinlabi 1996, Russell 1995). Possible surface realizations of such a floating feature often reach great, practically infinite numbers, so that without its abstract characterization the morpheme could not be distinguished from others at all. High-toned verbs in Chizigula, spoken in Tanzania, illustrate the case (cf. Kisseberth 1992). As in many Eastern Bantu

languages, high tone in Chizigula is almost never realized on the vowel it is morphologically associated with, but shifts further away from it, in some cases onto the next low-toned word or even onto the second following low-toned word within the scope of the phonological phrase. A possible mental image of the high-toned verb is thus its segmental make-up and a high tone appearing somewhere else. This abstract, discontinuous representation of the morpheme seems much more natural than the alternative of distinguishing different (i.e. low and high toned) “allomorphs” of all words which can incidentally be hosting the high tone of the verb. In chapter 7, I will present some evidence that language speakers carry mental images different from surface representations and may change their grammar according to them. With this one exception, throughout this work the issue of URs or I:O constraints will be of little importance to the discussion of analogy and all my analyses are compatible with either one-leveled or two-leveled versions of OT.

Returning to the problem of extended correspondence, I propose that a rational and sufficient way of constraining its open architecture consists in providing a motivation for each particular ranking of a given pair of Cor. In the large majority of cases this motivation comes from language usage and the frequency and semantic criteria distinguished above. Technically speaking, even though in theory Cor- $\{X:Y\}$  and Cor- $\{Y:X\}$  are equally possible, only one of these constraints may be licensed by language usage, which leads to its dominance over the other. Let us consider an uncomplicated hypothetical example of two related words  $X$  and  $Y$  with two possible different allomorphs  $a$  and  $a'$  of the same stem, conditioned by a phonological or morphophonological constraint, abbreviated as Alt. For simplicity, let us assume that there are no other words in the language with that particular stem in either allomorphic shape. If  $X$  and  $Y$  are members of an inflectional paradigm, each of them has its own affix ( $aff_1$  and  $aff_2$ ), so that the ultimate morphophonological structure of the words may look as in the left-hand column in (7a) below. Let us assume in addition that another pair of words  $W$  and  $Z$  were historically subject to the same alternation, by which their stem varied between  $b \sim b'$ , as in the left-hand column in (7b), but at a later stage  $b'$  was analogically replaced by  $b$ , due to its low frequency. If frequency criteria remained basically unchanged, they motivate a synchronic constraint ranking sketched in (7c).

(7)

	earlier stage	present stage	frequency
a/	X=a-aff <sub>1</sub> Y=a'-aff <sub>2</sub>	X=a-aff <sub>1</sub> Y=a'-aff <sub>2</sub>	high high
b/	W=b-aff <sub>1</sub> Z=b'-aff <sub>2</sub>	W=b-aff <sub>1</sub> Z=b-aff <sub>2</sub>	medium very low
c/	Cor- $\{Z:W\}$ >> Alt >> Cor- $\{X:Y\}$ , Cor- $\{Y:X\}$ , Cor- $\{W:Z\}$		

Since changes in usage are typically very slow, the correlation between usage and grammar as depicted above holds in the majority of cases. On rare occasions, the synchronic ranking appears meaningful only under a diachronic explanation, and in

extremely few cases, the ranking seems opposite to the expectation. Throughout this work, I will concentrate on the issue of constraint ranking motivation, believing that it provides the merit of the analysis, while a formal notation will be kept as simple as possible and restricted to illustrative examples only. I will now turn to the notion of the Base and more detailed description of Cor constraints viewed somewhat differently from the perspective of usage and grammar.

From the perspective of language usage, the notion of the Base in the present framework closely resembles Bybee’s (1985, 2001) lexical strength, which is directly correlated with frequency. In this sense, “Basehood accessibility” has a gradient character and increases with each occurrence of a given linguistic unit in language usage, so that it can better serve as a model for a non-Base, i.e. its correspondent. One possible approach to OT-style evaluation is directly from the usage viewpoint, i.e. by taking into consideration actual occurrences of particular forms. If, for example, a hypothetical word *W* from (7) above is fifty times more frequent than *Z*, its Base accessibility is fifty times stronger. A possible interpretation of this fact in the present framework can be that a potential word sharing the stem with *W* or *Z*, enters into 50 correspondence relations with *W* as the Base (stem *b*) and only one with *Z* as the Base (stem *b'*), by which the leveling to the latter stem is 50 times more costly. This is illustrated in (8a), where two hypothetical scenarios are compared, one with the *b* allomorph of word *W* leveled, and an opposite situation with the *b'* allomorph of word *Z* mapped onto *W*. Alternatively, we can interpret the disproportion in frequency as a need to evaluate as many as 50 *W* outputs for each *Z* output, as illustrated in (8b). This time Cor can be assumed to apply in a usual one-to-one fashion, hence is identically violated for each pair of outputs, and the decision making is left to Alt. The intuitive sense of the analysis in (8a) lies in the fact that the stem of the more frequent word is better stored, while the analysis in (8b) stresses usage conservatism.

(8a) Language usage evaluation (I)

	50xCor- $\{W=b\}$ , Cor- $\{Z=b'\}$	Alt
* $W=b\text{-aff}_1$ , $Z=b\text{-aff}_2$	*1	* (Z)
$W=b'\text{-aff}_1$ , $Z=b'\text{-aff}_2$	*50	* (W)

(8b) Language usage evaluation (II)

	Cor- $\{W=b\}$ , Cor- $\{Z=b'\}$	Alt
* 50 $W=b\text{-aff}_1$ , $Z=b\text{-aff}_2$	*	*1 (Z)
50 $W=b'\text{-aff}_1$ , $Z=b'\text{-aff}_2$	*	*50 (W)

Analyses such as those in (8) capture what can be called a “direct usage evaluation”, which, according to the model sketched earlier in (3), leads to, or “feeds” the grammar. I assume that in principle, the formulation of constraints as well as their ranking should follow from direct usage evaluation rather than being assumed by it. Hence, in the above example, the Base asymmetry in correspondence follows directly from frequency criteria. Its dominance over Alt (for ease of exposition already included in (8)) also follows from usage, specifically, from the token frequency threshold – more

frequent lexical items may dominate Alt, as do the hypothetical  $X$  and  $Y$  from (7) above. Likewise, the existence of a category of such alternating lexical items, i.e. type frequency, leads to the emergence of Alt as a constraint. To sum up, language usage data with their own evaluation provide substance for “grammar”, to which I will turn now.

I assume that detailed data regarding the frequency of particular linguistic units etc. are not explicitly incorporated into a more abstract, mental grammar, which has a more generalized and discrete character. Therefore, the quantity correspondence relations depicted in (8) will have a quality equivalent in an OT model, specifically, they will be represented by unidirectional correspondence constraints. Hence, the special status of the Base, which in (8) follows from simple majority criteria, must be expressed in “phonologized” terms. In a non-derivational theory such as OT, this presents a certain problem, because parallel evaluation may not apply to the Base in a different fashion than to its correspondent, and essentially such a result is desirable. In the OT literature, some attempts have been made in order to achieve the effect of a phonologically predicted Base, but none of them has been fully satisfactory. In my own earlier work, I have proposed a rather *ad hoc* constraint granting the Base a privileged status (Kraska-Szlenk 1995/2003, ch. 3). Recursive output evaluation has been put forward by Benua (1995, 1997), but this strategy can apply only to morphologically complex words (and it highly resembles derivational steps). The Base-less approach of McCarthy (2005) does not account for empirically attested facts, such as, for example, the Polish data discussed in chapters 2-4.

The above problems can be solved in the one-level model by the simple assumption that the Base in a correspondence constraint is the actual surface form, which also coincides with the input in two-level OT. This understanding of the Base complies with empirical and theoretical facts. Once underspecified inputs are eliminated from the grammar, at least one surface allomorph of a given stem must be stored in the speaker’s memory in order to retrieve a matching phonological form for a semantic concept. Such listed “input” can be formalized as a relatively high constraint, as first proposed by Russell (1995). Since the Base normally constitutes the most frequent allomorph, the simplest assumption is that it is listed, e.g. in the case of the Polish data of (5), we can assume  $\otimes \equiv [kf'at]$ . Therefore, using a surface Base for the purpose of correspondence does not involve any additional “cost” that the grammar would pay, because it is available in it any way. Whether all allomorphs are “listed” or only some of them, with others being derived by grammar, seems to be a question of little importance. “Listing” does not exclude being predictable by a rule (constraint), as convincingly argued in Langacker (1987) and many works since then. Listing and deriving are rather gradient and not categorical notions, which means that, with the exception of completely new linguistic units (e.g. made-up words, some proper names, novel metaphors etc.), all of them are to some extent listed and to some extent derived.

Returning to the issue of Cor constraints, we can now interpret the earlier example of Cor- $\{Z:W\}$  as “the stem of  $Z$  must be  $b$ ” (i.e. the “correct” stem of  $W$ ). Whether  $Z$  derives the  $b$  structure from this particular constraint or has it “listed” and only “checked” by the constraint does not have serious consequences for the grammar or language use. On the other hand, the dominated mirror-image Cor- $\{W:Z\}$  is quite invisible and can not be learned, which leads to the historical loss of the  $b'$  allomorph.

In real life, inflectional paradigms typically contain more than just two word-forms as assumed in the above hypothetical example, so that several word-forms may have one allomorph and several others – another one. The empirical evidence discussed in chapters 2-4 shows that analogical leveling takes place according to majority criteria computed over the entire paradigm, being directed towards the most frequent stem allomorph and not towards one particular word-form, e.g. the unmarked nominative (although the latter may play some role in rare cases of equal frequencies of different allomorphs, as I will suggest later in chapter 3). Therefore, the Base can be also understood as a stem abstracted from various word-forms and not necessarily a stem associated with a particular word, as I have assumed so far. These two versions of the Base are not in conflict with each other, but constitute a more concrete representation and a more schematic one, respectively (cf. Bybee 2001, Langacker 1987, 1991). However, in the following chapters, I will mostly use a word-to-word notation of correspondence constraints, since it helps to immediately identify the words that are crucial for establishing the Base.

In a similar fashion as correspondence constraints, Alt – a constraint predicting phonological or morphophonological alternation, emerges out of usage evaluation, with its relative ranking determined by quantity criteria.

Since an abstract grammar is to a large extent usage-driven, the Base typically coincides with the most frequent allomorph. Consequently, a question may be posed, whether usage evaluation should not suffice and simply constitute grammar? As already suggested earlier, the answer is “no”, because of possible discrepancies between the two, which necessitate a certain degree of arbitrariness in grammar. One such situation may result from a change in the frequency of a particular word, which makes the synchronic Cor ranking look unmotivated. Another example is found in already mentioned cases of similar frequencies of different allomorphs. In addition, there are changes in grammar leading to changes in usage, as in the case of hypercorrection or imperfect learning, when a Base is a “false” mental representation.

When usage and grammar evaluation predict different outputs as optimal, the discrepancy between them may initiate a change going in either direction. The fact that changes in usage lead to changes in grammar is quite straightforward and will be illustrated throughout this work. Perhaps less obvious, and also less frequent, are changes going in the opposite direction, but these will also be exemplified, especially in chapter 7.

#### **1.4. The data sources, research method and the study’s organization**

In this work, I concentrate on inflectional paradigms, since this is the area least understood from the theoretical perspective. The data come from various languages, but mostly from Polish declension (chapters 2-4). I will discuss three vowel alternation patterns, which vary greatly as to their current status as active constraints. A brief historical background will be provided to show that none of the processes is sufficiently motivated and transparent from the perspective of Polish synchronic phonology. But that does not impede the productivity of some of the alternations, which largely depends on the size of the lexical class being affected by them. In addition, Polish data are discussed in chapters 5 and 6, and *passim*.

To estimate the frequency of Polish lexical items and sizes of particular morphophonological categories, I have used two major sources, the frequency dictionary *Słownik* (1990) and the electronic Państwowe Wydawnictwo Naukowe (PWN) Corpus, which I will discuss in turn.

The frequency dictionary *Słownik* (1990) is based on a relatively small corpus of 500.000 words, divided into five components of 100.000 words each, representing five different genres. Each of these component parts consisted of 2.000 samples of small texts of about 50 words, randomly chosen from a large corpus of newspaper texts (two different genres), scholarly books and articles (including science and humanities), prose and drama (including radio plays). In the process of preparing the dictionary, some selection was performed to eliminate acronyms, quotations from foreign languages and the majority of proper names with the exception of names of nationalities and certain geographical names. The dictionary provides ranks and frequencies for each of the genres separately as well as for the whole corpus. In the present work, I will always use the latter option, i.e. for the entire 500.000 word corpus. The complete dictionary contains 10355 lexemes (with homonyms ranked separately) and does not include all words that appeared in the original corpus, but only those which occurred at least four times.

*Słownik* (1990) has certain drawbacks that should be pointed out. The dictionary is strongly biased towards the language of the 1960s in its journalistic style, since all the language material in the corpus (drawn from the authors' earlier work) was written in years 1963-1967 and 40% of it comes from newspapers. Another problem is that only written texts were included and merely 20% of the drama style imitates the spoken language. Therefore, we can see occasional paradoxes such as, for example, the noun *produkcja* 'production' ranked as high as the verb *iść* 'go', or common nouns, such as *bagaż* 'luggage', *obrus* 'table-cloth' or *pieróg* 'dumpling' ranked as low as some truly rare specialized words, e.g. *antygen* 'antigen', *carat* 'tsarism' or *Międzynarodówka* 'the Internationale'. In spite of these drawbacks, the dictionary turns out to be a useful tool, particularly for the purpose of estimating the size of a given class of words in a defined frequency range. In this work, I usually refer to three frequency ranges, namely, the first thousand most frequent words (ranks 1-1002 with 66 or more occurrences), the relatively frequent to medium vocabulary of the second thousand (ranks 1003-2009 with 65-32 occurrences) and the relatively rare vocabulary (four-occurrence words ranked 8739-10355).

The large electronic PWN Corpus contains 40.000.000 words and consists of fragments of various books, journals and newspapers, recorded conversations, as well as contents of web pages and advertisements. The corpus is balanced with respect to various genres and subject matters, but the spoken texts constitute only 4,5% of the total. Most of the language material comes from recent years 1990-2005 (78%); smaller samples of earlier texts date as far back as the year 1920. The corpus is available (for free in most of its functions) on the Internet on the following web page, <http://korpus.pwn.pl>, where more information can be found as to its contents and organization.

I have used the PWN Corpus in its full version for the purpose of estimating frequency of use of particular WORDS and their word-forms, whenever I thought the *Słownik* (1990) data were too rough, as well as in cases of rare words, not included in the dictionary. Although in general the PWN Corpus appears to be a very convenient and reliable source, in some cases its use was limited, due to the fact that it does not

disambiguate homonyms. Thus, for example, the frequency of the noun *siano* ‘hay’ can not be easily determined, because some of its word-forms are homonymic with the word-forms of the verb *siać* ‘sow’, cf. *siano* ‘hay-nom.’ or ‘it was sowed’, *sianie* ‘hay-loc.’ or ‘sowing-nom.’, etc. Homonymic pairs of this kind are not numerous, but several nouns had to be excluded as “untestable”, especially in chapter 2.

Even though I always specifically mention which of the above two sources is meant, I feel obliged to caution the reader against a possible confusion resulting from my referring to both of them in the same work, often in the same section (but probably not on the same page). Since the PWN Corpus is eighty times larger than *Słownik* (1990), a word judged as “very frequent” by the former would have thousands of occurrences – many times more (ideally eighty times more, if the two were perfectly equivalent) than in the latter source with several hundreds of occurrences only. For example, the lexeme *ŚWIAT* ‘world’ has 29290 PWN Corpus occurrences, which is 75 times more than the 389 occurrences of this word in *Słownik* (1990). On the other hand, a relatively rare word could have about a hundred occurrences in the PWN Corpus, while such a number of occurrences is characteristic of very frequent words listed in *Słownik* (1990). I hope this explanation should prevent the reader from comparing absolute values of figures coming from the two different corpora.

As to historical sources, I have used mostly Długosz-Kurczabowa (1998) and Rospond (2003), as well as the etymological dictionaries of Bańkowski (2000), Brückner (1974) and Sławski (1952-1974). A lot of detailed information concerning Old Polish has been drawn from the comprehensive 32-volume dictionary of 16<sup>th</sup> century Polish (*Słownik* 1968-2004). This source, compiled on the basis of a corpus of 200 sampled texts, has been particularly useful, since it contains the statistics of occurrence of WORDs and word-forms in the corpus. Reference to these and other historical sources is made explicitly only when specific information provided by the given author is cited and not for commonly known facts.

I use rough IPA transcription of the Polish data leaving aside certain phonetic details. In particular, I mark the phonemic distinction between [e] and [ɛ] in the data of Early Polish, but I do not distinguish allophonic and sometimes gradient distinctions between these two vowels in the data of the contemporary language using only [e]. Likewise, the vowel *o* is uniformly transcribed as [ɔ], although in rare contexts it approximates [o]. In certain cases, only Polish orthography is used (with English glosses), especially in presenting word-lists (sometimes very long), since the only concern is whether the words do or do not undergo a given alternation and this is explicitly stated. The following rules of Polish orthography should be helpful in reading the data.

Vowels are straightforward, except for the following: *a* and *e* are read as [ɔN] or [ɔw̃] and [eN] or [ew̃], respectively; *ó*=u=[u], *y*=[i]; *i* after a consonant (and before a vowel) indicates its palatalization, as in e.g. *siano* [ɕano] ‘hay’. Word-finally and before a consonant, palatalization is indicated by a diacritic, e.g. *ś*=[ɕ], *ń*=[ɲ]. Consonants are straightforward except for: *c*=[ts], *ch*=*h*=[x], *cz*=[tʃ], *dz*=[dʒ], *dź*=[dʒ̣], *j*=[j], *ł*=[w], *rz*=*ż*=[ʒ], *sz*=[ʃ].

As to the other languages treated here, Standard Swahili data are discussed at some length in chapter 6 and *passim*. For the presentation of the basic facts, I have used my own knowledge of the language, which I have been studying and teaching for many years. More specific data come from the sources indicated in the text. Frequency of use was estimated with the help of the electronic Helsinki Corpus of Swahili, which was made available to me by kind permission of Arvi Hurskeinen.

The data concerning the basic patterns of Classical Arabic, as well as more detailed examples from modern Arabic dialects are discussed in chapter 5. I do not give reference for the former, using my own knowledge of the language, refreshed with the help of standard grammars and dictionaries. A number of sources, indicated in the text, have been used for the latter.

English stress data are briefly presented in chapter 8 on the basis of the literature of the subject cited therein. The British National Corpus in the form available on the Internet served as a source of frequency counts.

Examples from other languages include Yiddish (chapter 7), Romance (chapter 6) and others, and are quoted from indicated sources.

The theoretical issues discussed in the remaining chapters of this work are organized as follows. A case study of Polish declensional patterns in chapters 2-4 provides empirical evidence for the role of frequency in analogy, from the synchronic, as well as the diachronic perspective. The recessive *e~a/o* alternation is discussed in chapter 2, the stable *o~u* alternation in chapter 3, and the most productive *e~∅* alternation in chapter 4. Chapter 5 discusses the correlation between analogy and semantic distance, which leads to asymmetries between nouns and verbs, as well as between inflection and derivation. The two following chapters concern cases of ‘pseudo-analogy’, when stem leveling comes as a by-product rather than a goal itself, due to such factors as the size of a linguistic unit vis-à-vis its frequency (“Zipf’s laws”, ch. 6) and hypercorrection, or “UR-driven” analogy (ch. 7). The final chapter 8 contains some comments on other theoretical issues, e.g. the problem of analogy in derivational morphology or the implication of the present model for understanding of the lexicon-grammar connection.

## CHAPTER 2

# The *e~a/o* alternation in Polish nouns

### 2.1. Preliminaries

The alternation discussed in this chapter goes back to the historical process called the Lechitic Vowel Shift, since it affected all Lechitic dialects, i.e. Polish, Pomorian and Polabian, but not the closely related Czech language, which was already separated and which remained untouched by the innovation. The shift changed front vowels {*e*, *ɛ*, *ě*} and soft syllabic sonorants {*l̥*, *r̥*} before non-palatalized coronals {*t*, *d*, *n*, *s*, *z*, *r*, *l*} to the back variants {*o*, *a*, *ō*, *l̥*, *r̥*}, respectively. The process is documented in the 9<sup>th</sup> and 10<sup>th</sup> century chronicles (containing many Polish proper names) and was fully active at the time of the adaptation of early loanwords relating to the Christian religion accepted by Poland in 966 from the Czechs, cf. *anioł* < Cz. *anjel* < Lat. *angelus* ‘angel’; *kościół* < Cz. *kostel* < Lat. *castellum* ‘church’; *ofiara* < Cz. *ofer* < Ger. *Opfer* ‘sacrifice’. It is commonly agreed that the shift was no longer productive by the early 12<sup>th</sup> century, since it did not affect the *e* vowel newly developed from the so-called “yers” (cf. chapter 4).

The change had some consequences for the Polish phonological system, which previously harmonized consonants and vowels in [backness]: front vowels followed only palatalized (“soft”) consonants and back vowels followed non-palatalized (“hard”) consonants. After the shift, palatalized consonants acquired the phonemic status due to their occurrence before the back *a* or *o*, cf. *las* ‘forest’, *kwiat* [kʲfat] ‘flower’, *siodło* [ɕodwɔ] ‘saddle’, *miotła* [mʲotwɔ] ‘broom’ etc. More importantly, the shift introduced alternations of the stem vowel in inflection and derivation which still occur in the contemporary language, unless they were eliminated by analogy. I will concentrate on a larger class of the *e~a/o* cases, excluding from the discussion a handful of rather irregular data involving the development of syllabic sonorants, as well as the original alternation of nasal vowels, which was very much obscured by subsequent changes that affected them.

The *e~a/o* alternation is found in words of various categories: nouns, verbs, adjectives. It may induce stem allomorphy intraparadigmatically, cf. *las* ‘forest-nom.’, *lesie* ‘forest-loc.’, or across categories, cf. *biały* ‘white’, *bielić* ‘to whiten’. In this chapter, I will focus on nominal paradigms only; a brief discussion of verbs will be included in chapter 5.

As illustrated by the examples below, in nouns of all three genders, the palatalizing environment occurs in the singular locative case for which the *e*-variant is predicted. The same form is also used as the dative of feminine nouns and as the vocative of masculine nouns. In addition, the palatalizing environment occurs in the nominative=vocative plural of those masculine nouns denoting persons that have the *-i* case ending instead of the more common *-owie*. Other than that, all forms in the singular and the plural have back *o* or *a* stem variants, which therefore constitute a major pattern

as far as the number of cases in the paradigm is concerned.<sup>6</sup> The complete paradigms of *a~e* alternating nouns of all three genders are shown in (9).<sup>7</sup> The *o~e* alternation is found in four nouns (three stems) of the masculine gender only, the two listed in (10) as well as *popiół* ‘ash’ and *archanioł* ‘archangel’. The minor *e*-variant forms are underlined.

(9a) The singular paradigm of *a~e* nouns

case	masculine	feminine	neuter
nom sg	las, sąsiad	gwiazda	miasto
gen sg	lasu, sąsiada	gwiazdy	miasta
dat sg	lasowi, sąsiadowi	<u>gwieździe</u>	miastu
acc sg	las, sąsiada	gwiazdę	miasto
instr sg	lasem, sąsiadem	gwiazdą	miastem
loc sg	<u>lesie</u> , <u>sąsiedzie</u>	<u>gwieździe</u>	<u>mieście</u>
voc sg	<u>lesie</u> , <u>sąsiedzie</u> ‘forest’, ‘neighbor’	gwiazdo ‘star’	miasto ‘town’

(9b) The plural paradigm of *a~e* nouns

case	masculine	feminine	neuter
nom pl	lasy, <u>sąsiedzi</u>	gwiazdy	miasta
gen pl	lasów, sąsiadów	gwiazd	miast
dat pl	lasom, sąsiadom	gwiazdom	miastom
acc pl	lasy, sąsiadów	gwiazdy	miasta
instr pl	lasami, sąsiadami	gwiazdami	miastami
loc pl	lasach, sąsiadach	gwiazdach	miastach
voc pl	lasy, <u>sąsiedzi</u> ‘forests’, ‘neighbors’	gwiazdy ‘stars’	miasta ‘towns’

<sup>6</sup> In Old Polish, there was one more case ending with the palatalizing environment, pl. loc. *-ech*, which was replaced with the back vowel *-och* in the 14<sup>th</sup> c. and later on with the present suffix *-ach*. Given that the new suffix appeared relatively shortly after the vowel shift, and given the very low frequency of the pl. loc. case, I will ignore this fact in subsequent discussion.

<sup>7</sup> Throughout this work, for the reader’s convenience, I present the data in a uniform manner and indicate the same three genders in the singular and the plural. It should be noted, however, that for morphosyntactic reasons, there are only two genders in plural: the so-called “masculine-personal” (since it covers masculine nouns denoting people only) and “common” or “non-masculine-personal”, which puts together feminine, neuter and non-personal masculine nouns.

(10) The paradigm of a *o~e* nouns

case	singular	plural
nom	anioł, kościół	aniołowie/ <u>anieli</u> , kościoły
gen	anioła, kościoła	aniołów, kościołów
dat	aniołowi, kościołowi	aniołom, kościołom
acc	anioła, kościół	anioły, kościoły
instr	aniołem, kościołem	aniołami, kościołami
loc	<u>aniele</u> , <u>kościel</u>	aniołach, kościołach
voc	<u>aniele</u> , <u>kościel</u> 'angel', 'church'	anioły, kościoły 'angels', 'churches'

It is not only that the *e*-variant creates a minor pattern within the paradigm, but it also turns up as relatively rare as far as text frequency is concerned. Generally speaking, the most frequent cases in Polish appear to be the nominative, the genitive and the accusative (not necessarily in that order), none of which has an *e*-form. Dative singular, which is relevant for the feminine declension, is mostly used in the benefactive meaning (with verbs as *dać* 'give' etc.), which is practically limited to animate nouns. Similarly, the vocative case, relevant for the masculine paradigm, is normally used only with people's names and a few other forms of address. Among the data discussed here, there are two examples for which the vocative is of significant use. One is the noun from table (9a) above, *sąsiedzie* 'neighbor', which can be used as a semi-informal form of address. The other one is *aniele* 'angel', shown in (10), occurring in common Catholic prayers, which can be also used in a joking manner to address a loved one. The next inflectional case, nominative plural, is relevant for the same two nouns (and for no others, as far as I can tell). Contrary to the general tendency for higher frequency of singular than plural, *sąsiedzi* 'neighbors' occurs slightly more often than its singular counterpart *sąsiad* 'neighbor' (442 times to 317 in the PWN corpus data), which considerably increases the mean occurrence of the minor variant for this noun. The plural *anieli* 'angels' is used in prayers and Christmas carols, even though nowadays is subsequently replaced with the alternative form *aniołowie*. Finally, the locative singular, which has the *e*-allomorph in nouns of all three genders, is used mostly with prepositions expressing static locations, as *w* 'in' or *na* 'on' (as well as with some other prepositions of limited use, e.g. *o* 'about', as in *mówić o* 'to talk about'). With the exception of nouns denoting places, which occur in the locative quite often<sup>8</sup>, the average frequency of that case is low. This claim will be soon supported by the actual data from the corpus. In the meantime, it is worthwhile to observe that the low frequency of the minor *e*-variant creates a perfect opportunity for this form to be analogically replaced with the major *a/o* pattern, and not vice versa. Indeed, this will be the observable result in the great majority of cases.

First, however, let us consider the distribution of nouns exhibiting the alternation, as well as those leveled by analogy. I have used the frequency dictionary (*Słownik* 1990) to estimate the size and the relative frequency of the class of nouns exhibiting the alternation, as well as those which were leveled by analogy.

---

<sup>8</sup> This is especially true of proper names of places, which often take the locative case as a base for analogy (cf. Mańczak 1996: 217).

## 2.2. The distribution of alternating and leveled nouns at different frequency ranges

Within the most frequent vocabulary, ranked from 1 to 1002, ten alternating nouns are found, shown with their ranks and frequency in the table below.

Table I. The *e~a/o* alternating nouns ordered according to their rank, as they appear in the first 1000-word list (with the gender marked in parentheses).

noun (nom.) + gloss	rank	frequency
miasto (n) 'town'	101	452
świat (m) 'world'	126	389
ciało (n) 'body'	364-366	159
światło (n) 'light'	364-366	159
powiat (m) 'district'	522-527	116
kościół (m) 'church'	557-564	110
las (m) 'forest'	565-578	109
miara (f) 'measure'	693-704	91
wiatr (m) 'wind'	836-845	77
zjazd (m) 'reunion'	959-981	67

Within the same range, there are five nouns with the *a/o* vowel analogically leveled throughout the paradigm, listed below. In the case of *czoło*, the earlier *e*-variant of the locative is still used in the fixed expression *na czele (pochodu etc.)* 'in the forefront (of a demonstration etc.)' on a par with the new analogical form *na czole* 'on the forehead'. For all these nouns there exist morphological derivatives with the *e*-variant of the stem, cf. *wczesny* 'early', *żeński* 'female', *oddzielny* 'separate', *ścienny* 'wall (adj.)', *naczelnik* 'leader'. There is also one noun (the underlined *cena* 'price') with the opposite direction of leveling, i.e. towards the locative/dative *e*-variant. The older form *\*cana* is attested historically.

Table II. Nouns in the first 1000-word list with the environment for the *e~a/o* alternation with fixed vowels due to analogy.

noun (nom.) + gloss	rank	frequency
czas (m) 'time'	51	848
żona (f) 'wife'	485-490	123
<u>cena</u> (f) 'price'	530-540	114
oddział (m) 'department'	639-649	98
ściana (f) 'wall'	652-661	96
czoło (n) 'forehead'	736-749	86

In the second thousand of the most frequent vocabulary, there are five alternating nouns and eight non-alternating, leveled by analogy, as shown in the tables III and IV,

respectively. With the exception of *dzieło* ‘work’, which is in fact a case of a lexical split, all nouns are leveled towards the major *a~o* stem pattern.

Table III. The *e~a/o* alternating nouns ranked between 1003-2009.

noun (nom.) + gloss	rank	frequency
kwiat (m) ‘flower’	1165-1185	55
wiara (f) ‘faith’	1414-1443	45
gwiazda (f) ‘star’	1444-1475	44
obiad (m) ‘dinner’	1700-1757	37
sąsiad (m) ‘neighbor’	1910-1941	33

Table IV. The *e~a/o* nouns leveled by analogy, ranked between 1003-2009.

noun (nom.) + gloss	rank	frequency
<u>dzieło</u> (n) ‘work, composition’	1003-1018	65
ślad (m) ‘trace’	1271-1301	50
dział (m) ‘department’	1444-1475	44
lód (m) ‘ice’	1541-1570	41
siostra (f) ‘sister’	1541-1570	41
żelazo (n) ‘iron’	1698-1757	37
jezioro (n) ‘lake’	1806-1853	35
wiosna (f) ‘spring’	1942-2008	32

Throughout the subsequent medium and lower ranks, the proportion of alternating nouns considerably diminishes with respect to the number of those leveled by analogy. In any case however, the joint number of both kinds of nouns, i.e. including those which were historically alternating or, are potentially “alternable” from the synchronic perspective, continues to be low. This is precisely a factor favoring stem-analogy to the exclusion of pattern-analogy. The latter option seems to be particularly discouraged in the case at hand also because of the additional fact that this group of nouns is rather unproductive and lacks common morphological or semantic features which could unify it as a salient morphophonemic class. The large majority of the nouns have monomorphemic stems or contain a frozen unrecognizable suffix (cf. *-lo* as in *wiosło* ‘paddle’), and even if some of them contain a productive derivational suffix, it does not recur in this particular group of nouns (cf. *-adło* as in *zwierciadło* ‘mirror’). Another factor enhancing stem-analogy is the existence of numerous common nouns that entered the lexicon after the alternation became unproductive (12<sup>th</sup> century) and whose vowel is fixed in spite of the presence of the favorable environment, cf. *bocian* ‘stork’, uncertain etymology, 14<sup>th</sup> c., *los* (m) ‘lot; fortune’, from German, 14<sup>th</sup> c., *kobieta* ‘woman’, uncertain etymology, 16<sup>th</sup> c., *cera* ‘wax (arch.); complexion’, from Latin, 16<sup>th</sup> c., etc.

On the basis of the data collected from the contemporary dictionaries, Krzyżanowski (1983, 1992) finds fewer than thirty *e~a/o* alternating nouns in the Polish lexicon. The same author mentions as many as 115 non-alternating nouns, but since he

does not list them, it is hard to estimate the exact number of those which alternated in the past and were leveled.

The following table V shows the alternating nouns spotted in the remaining part of the frequency dictionary (between ranks 2001-10355). Table VI contains the nouns with the analogically leveled fixed vowel<sup>9</sup>, including one case of variation (*przód* ‘front’), one case of leveling to the minor locative (*krzesło* ‘chair’) and one case of a lexical split (*bieda* ‘poverty’). (Nouns with the leveled minor *e*-variant are underlined.)

Table V. The *e~a/o* alternating nouns ranked between 2001-10355.

noun (nom.) + gloss	rank	frequency
lato (n) ‘summer’	2662-2767	22
gniazdo (n) ‘nest’	3554-3754	15
popiół (m) ‘ash’	5147-5583	9
anioł (m) ‘angel’	5584-6133	8
ciasto (n) ‘cake’	7623-8738	5

Table VI. The *e~a/o* nouns leveled by analogy, ranked between 2001-10355.

noun (nom.) + gloss	rank	frequency
przód (m) ‘front’	2010-2069	31 (loc. <i>przodzie~przedzie</i> )
<u>krzesło</u> (n) ‘chair’	2191-2262	28
<u>bieda</u> (f) ‘poverty’	4217-4486	12
siano (n) ‘hay’	4786-5146	10
miód (m) ‘honey’	5584-6133	8
dziad (m) ‘forefather; pauper’	6134-6804	7
zwierciadło (n) ‘mirror’	6134-6804	7
wiosło (n) ‘paddle’	7623-8738	5
wiadro (n) ‘bucket’	8739-10355	4
wrzód (m) ‘ulcer’	8739-10355	4

The data discussed thus far suggest that there seems to be a correlation between the frequency of the word and its susceptibility to stem analogy. In order to examine this relation more carefully, I will compare the frequencies of particular WORDS and their word-forms using the larger electronic PWN corpus. For the purpose of this test, I will use the nouns included in the *Słownik* (1990), as well as additional lexemes of lower frequency.

The data in table VII below show detailed frequencies of particular word-forms within the paradigms of the most frequent alternating nouns – those which appear more

<sup>9</sup> I included only the nouns for which the alternating form is without doubt historically attested. I excluded all deverbal nouns, whose stem never alternates (even though it does in verbal inflection, cf. *lot/locie* ‘flight nom./loc.’ vs. the verb *leci* ‘flies-3 sg.’), because they may be treated as a sub-paradigm of their own (and often a morphological derivative is recent enough not to be subject to the historical alternation any way).

than 5000 times in the corpus, ordered according to their decreasing frequency.<sup>10</sup> Joint frequencies are also noted, as well as the percentage of the minor *e*-variant. It is worthwhile to observe that out of the seven words, four are names of places for which the locative case (the *e*-variant) is frequently used: it is the second most frequent word-form for *świat* ‘world’, the third one for *miasto* ‘town’ and *kościół* ‘church’, and the fourth most frequent word-form for *las* ‘forest’. The locative has also very high frequency in the case of *światło* ‘light’, because of the common phrase *w świetle* ‘in the light [of]’. Even though for the remaining two words, *ciało* ‘body’ and *ofiara* ‘sacrifice’, the percentage use of the locative is relatively small, its actual occurrence in the corpus is rather high (almost 500 times and over 100 times, respectively).

Table VII: The frequencies of word-forms of alternating nouns occurring over 5000 times in the PWN corpus.

ŚWIAT ‘world’ 29290: *świata* 11600, *świecie* 9161 (**31,28%**), *świat* 6607, *światem* 1032, *światu* 499, *światy* 181, *światów* 153, *światach* 29, *światami* 22;  
 MIASTO ‘town’ 17930: *miasta* 7478, *miasto* 3202, *mieście* 2845 (**15,87%**), *miast* 2115, *miastach* 1244, *miastem* 712, *miastu* 208, *miastami* 93, *miastom* 24;  
 KOŚCIÓŁ ‘church’ 9656: *kościół* 3604, *kościół* 3053, *kościelnie* 1239 (**12,83**), *kościółem* 493, *kościółów* 486, *kościóły* 320, *kościółach* 197, *kościółowi* 175, *kościółami* 70, *kościółom* 19;  
 CIAŁO ‘body’ 7232: *ciała* 3392, *ciało* 2163, *ciałem* 507, *ciał* 511, *cielenie* 496 (**6,86%**), *ciałach* 48, *ciału* 53, *ciałami* 44, *ciałom* 18;  
 OFIARA ‘sacrifice’ 5803: *ofiar* 2004, *ofiary* 1374, *ofiarą* 809, *ofiara* 429, *ofiarami* 377, *ofiare* 348, *ofiarom* 262, *ofierze* 127 (**2,19%**), *ofiarach* 69.  
 ŚWIATŁO ‘light’ 5573: *światła* 1753, *światło* 1670, *świetle* 1402 (**25,16%**), *światłem* 342, *światel* 269, *światłami* 62, *światłach* 53, *światłu* 21, *światłom* 0;  
 LAS ‘forest’ 5465: *las* 1025, *lasu* 1173, *lasów* 973, *lesie* 837 (**15,32%**), *lasy* 658, *lasach* 409, *lasem* 223, *lasami* 124, *lasom* 18, *lasowi* 6.

Table VIII shows the figures for the WORD and the minor *e*-form of the medium frequency alternating nouns (including three cases of variation). At first glance, the data seem rather inconsistent, since the percentage and absolute values of the minor forms vary greatly from very high (e.g. *miara* ‘measure’, *sąsiad* ‘neighbor’) to very low (e.g. *gwiazda* ‘star’, *kwiat* ‘flower’). But it is worthwhile to point out that there are no truly low frequency nouns in this group – the lowest figure for WORD is 454 (in the case of ‘ash’). Likewise, the frequency of the minor form is never zero – the lowest figure being 20 (in the case of ‘angel’). An additional factor supporting the alternation in lower frequency words of this group is the presence of derivatives – I will return to this issue later in this chapter.

<sup>10</sup> The order determined on the basis of the electronic corpus may sometimes slightly differ from the ranking in *Słownik* (1990) – see the discussion in section 1.4.

Table VIII. The frequencies of the WORD and the minor *e*-form of alternating nouns occurring less than 5000 times in the PWN corpus.

MIARA ‘measure’ 4714: mierze 1369 (29,04%) ~ miarze 1 (0,02%)  
 WIARA ‘faith’ 3813: wierze 356 (9,34%)  
 GWIAZDA ‘star’ 3773: gwieździe 35 (0,9%)  
 CZOŁO ‘forehead’ 3365: czele 1609 (47,82%) ~ czole 255 (7,58%)  
 WIATR ‘wind’ 3328: wietrze 176 (5,29%)  
 POWIAT ‘district’ 2206: powiecie 328 (14,87%)  
 KWIAT ‘flower’ 2552: kwiecie 27 (1,1%)  
 SAŚIAD ‘neighbor’ 2355: sąsiedzie 17 + sąsiedzi 459 (20,21%)  
 ZJAZD ‘reunion’ 1635: zjeździe 305 (18,65%)  
 OBIAD ‘dinner’ 1630: obiedzie 265 (16,26%)  
 PRZÓD 1454: przedzie 39 (2,68%) ~ przodzie 59 (4,06%)  
 ANIOŁ ‘angel’ 985: aniele 20 + anieli 0 (2,03%)  
 CIASTO ‘cake, dough’ 815: cięście 48 (5,89%)  
 GNIAZDO ‘nest’ 770: gnieździe 62 (8,05%)  
 POPIÓŁ ‘ash’ 454: popiele 34 (7,49%)

The following table IX presents the nouns in which the minor *e*-form was leveled to the major *a/o*-stem form. Similarly to the previous case, if we look at the group as a whole, there seems to be only a small degree of correlation between the fact of leveling and frequency of particular word-forms. The nouns at the top of the list have very high frequencies of both the WORD and the leveled minor form. The nouns in the middle do not differ much in their frequencies from some of the nouns of the alternating class of table VIII. But it is only in this group that we find nouns whose WORD frequency is lower than 400 and it is only in this group that we find zero occurrences of the (historical) *e*-form. Most of the nouns comprising the lower part of the table combine low frequency of WORD with low percentage of the leveled form.

Table IX. The frequencies of the WORD and the minor *e*-form leveled to *a/o* in non-alternating nouns in the PWN corpus.

CZAS ‘time’ 58572: czasie 15193 (25,94%)  
 ŻONA ‘wife’ 6704: żonie 388 (5,8%)  
 ODDZIAŁ ‘department’ 5402: oddziale 549 (10,16%)  
 ŚCIANA ‘wall’ 4951: ścianie 786 (15,88%)  
 ŚLAD ‘trace’ 3543: śladzie 10 (0,3%)  
 JEZIORO ‘lake’ 2966: jeziorze 193 (6,51%)  
 SIOSTRA ‘sister’ 2631: siostrze 87 (3,3%)  
 LÓD ‘ice’ 2267: lodzie 309 (13,63%)  
 WIOSNA ‘spring’ 2230: wiośnie 40 (1,79%)  
 DZIAD ‘forefather (arch.); pauper’ 744: dziadzie 0 (0%)  
 MIÓD ‘honey’ 638: miodzie 27 (4,23%)  
 ZWIERCIADŁO ‘mirror’ 302 : zwierciadle 65 (21,52%)  
 BRZOZA ‘birch’ 243: brzozie 9 (3,70%)

SIODŁO ‘saddle’ 248: siodle 61 (24,60)  
 BIESIADA ‘feast’ 227: biesiadzie 15 (6,61%)  
 WIADRO ‘bucket’ 219: wiadrze 9 (4,11%)  
 WIOSŁO ‘paddle’ 157: wiośle 0 (0%)  
 WRZÓD ‘ulcer’ 147: wrzodzie 0 (0%)  
 MIOTŁA ‘broom’ 113: miotle 21 (18,58%)

The conclusion of the preceding discussion could be that the language usage approach can explain some of the data, especially at both ends of the frequency continuum, but not all of them. Many of the nouns in the higher and mid frequencies do not seem to follow any principle in being conservative and maintaining the alternation or being progressive and eliminating the alternation by analogy.

For a few cases here, it could be argued that the corpus does not reflect the real language use data, because some forms are more frequent in the spoken language than in written texts, which the corpus is mostly made of. For example, the last word in table IX has a relatively high percentage of the locative=dative because of the phrase *na miotle* ‘on the broom’ used in the corpus exclusively in the context of witches’ flying, which belongs more to the domain of literature than real life. At the same time, everyday expressions (with nouns of the major *o*-form), such as *gdzie jest miotła* ‘where is the broom’, *daj mi/weź miotłę* ‘give me/take the broom’ are underrepresented in the corpus.<sup>11</sup> Conversely, the locatives of nouns such as *obiedzie* ‘dinner’ or *cieście* ‘cake, dough’ may have higher frequency in the spoken language, because the former is used in common phrases as *na obiedzie* ‘during dinner’ or *po obiedzie* ‘after dinner’, coll. ‘afternoon’, and the latter is a component of compounds naming various kinds of deep-fried or baked dishes. It seems, however, that neither for these particular nouns nor for other examples would such readjustments significantly influence the results (as they do not for *miotle*, *obiedzie*, *cieście*, cf. tables VIII and IX). However, the frequency approach is capable of explaining the data, if additional factors are taken into consideration. I will now discuss them in turn.

### 2.3. Other factors enhancing/preventing analogy

Frequency of use of a given lexeme or a particular word-form may considerably change with time. Analogical leveling may affect a particular word when it has a limited scope of use, while at a later time the word may expand its usage due to a meaning extension or generalization, or a cultural or any other unpredictable reason. Once an alternation is eliminated from an unproductive pattern, it will not reappear, even though it would be tolerated well at the present stage. The top word in table IX, *czas* ‘time’ provides a perfect example of such a case. The present high frequency of the lexeme (58572) is due to its very general meaning; the present high frequency of the locative (25,94%) is mostly due to the common phrase *w czasie* ‘in the time [of], during’. Originally, the meaning of the noun was much narrower ‘defined time, due-date’. In Old Polish, the locative was used in infrequent phrases as e.g. *po czasie* ‘after due-time, past the dead-line’ (which is still used in the contemporary language) or *na czasie* ‘pregnant’ (no longer used in this

<sup>11</sup> Of course now, in the age of vacuum-cleaners and Harry Potters, the frequency data may get reversed (but that should not bother us, since analogy has already done its job).

sense). The meaning extension was not at first accompanied by the increase of the use of the locative, because other inflectional forms, *czasu* and *czas* (with the back variant of the stem), were used in the sense of ‘during’. The dictionary of 16<sup>th</sup> century Polish provides numerous examples, such as *czasu wieczornego* ‘in the evening time’, *czasu wesela* ‘in the time of happiness’, *czasu wojny* ‘during the war’, *czasu wiosny/w czas wiosny* ‘during springtime’, *w niebezpieczny czas* ‘in the time of danger’, etc. (*Słownik* 1969:26-104). In total, the examples of expressions with the lexeme *czas* occurring with the back stem variant encompass almost all 78 finely printed pages of the entry *CZAS* in this dictionary and constitute 98% of all corpus data (17252 occurrences). Out of the 350 occurrences (2%) of the locative, most (328) already have the analogical form *czasie*, while much fewer (22) – the original form *czesie*. The extreme rarity of use of the locative in the 16<sup>th</sup> century explains why it had already undergone analogy by that time.

In the case of the *e~a/o* alternation, stem leveling is a constant process affecting particular lexemes as their frequency decreases. This often happens when the object denoted by the given word begins to lose its importance in the life of the society or when a given lexeme is pressured by a synonym. Several low frequency lexemes from table IX associated with the rural life domain represent clear examples of the former case, e.g. *siodło* ‘saddle’, *wiadro* ‘bucket’, *miotła* ‘broom’, *wiosło* ‘paddle’, *brzoza* ‘birch’. Another example from that table, *zwierciadło* ‘mirror’, illustrates the latter case, which I will discuss at some length below.

The word *lustro* was borrowed into Polish in the 18<sup>th</sup> century and it denoted a wall candelabrum with reflecting mirrors. Gradually it extended its meaning and started to be used in the sense of ‘mirror’, replacing the native lexeme *zwierciadło*. Nowadays, the latter is completely eliminated from everyday use, although it still appears in metaphorical expressions, e.g. *zwierciadło sprawiedliwości* ‘the mirror of justice’ and in various less than “usual” contexts (e.g. it is appropriate for Snow White’s stepmother to use *zwierciadło* in the fairy tale, but it would sound very bizarre for someone to use the word to refer to a mirror, however beautiful it could be, in his/her house). The grammar’s response to the severe decrease of the word’s frequency was a gradual replacement of the locative *zwierciadle* by the analogical form *zwierciadle*. J. N. Baudouin de Courtenay in 1904 (Baudouin de Courtenay 1974:335), but also S. Rospond in 1969 (Rospond 2003: 56) cite both forms as free variants. At present, *zwierciadle* is not possible anymore (NB no occurrence in the PWN Corpus), only the analogical form *zwierciadle* is.

Let us now turn to the words in tables VIII and IX, which have similar frequencies, but behave differently with respect to analogy. What we can observe is a correlation between maintaining the alternation and the presence of derivatives based on the *e*-variant of the given stem. Most of the alternating words in table VIII (including all for which the *e*-word-form is not very salient) have semantically close derivatives. For example, the *e*-stem of *kwiat* ‘flower’ occurs in numerous derivatives, as e.g. *kwiecisty* ‘flowery’, *ukwiecić* ‘decorate with flowers’, *kwietnik* ‘flower bed’, etc.; of *wiara* ‘faith’ – in *wierzyć* ‘to believe’, *wierny* ‘faithful’, *wierność* ‘faithfulness’, etc.; of *anioł* ‘angel’ – in the adjective *anielski* ‘angel-like, i.e. good, calm’; of *gwiazda* ‘star’ in the adjective *gwiazdny* ‘star’; of *popiół* ‘ash’ – in the adjective *popielaty* ‘gray’; of *obiad* ‘dinner’ – in the adjective *poobiedni* ‘after dinner’, of *sąsiad* ‘neighbor’ – in the adjective *sąsiedzki* ‘neighboring’, etc. All of these derivatives are semantically transparent and it can be claimed that they support the alternation in the paradigms of the basic nouns. All

occurrences of the front variant of the given stem contribute to the total frequency of this allomorph, increasing its salience. Consequently, the pressure for analogical leveling within the paradigm is smaller than in the cases where such derivatives are absent. In fact, if a minor word-form of a basic noun has extremely small frequency, we can hypothesize that the “support” by more frequent derivatives is nothing else but an analogical process directed *towards* a derived form.<sup>12</sup> It might be easier for a speaker to retrieve a salient existing stem allomorph and map it to the same phonological but different morphological environment than to create a new morphophonological pattern (i.e. back stem+front suffix). Let me illustrate this with an example of the stem of *kwiat* ‘flower’. It has two salient allomorphs: the back variant [kfʲat], occurring in various word-forms of the noun ‘flower’, and [kfʲetɕ], occurring in “soft” environments in derivatives. A non-salient locative of the noun ‘flower’ can be analogically derived either from the first one, eliminating alternation within the paradigm, or from the second one, preserving the alternation. Although the former strategy may seem more natural, it creates a new allomorph [kfʲatɕ], since the coronal must undergo the obligatory palatalization before the *e* of the locative suffix.<sup>13</sup> Even though the difference between [kfʲat] and [kfʲatɕ] is minimal, it is a difference. Therefore, in the particular case here, it might be less costly to follow the latter option and this is what actually happens. Nonetheless, it should be noted that the choices the grammar makes in cases such as this are very subtle. In other words, it should not be surprising if paradigmatic leveling actually would affect the locative of ‘flower’ in the near future. Strikingly, as already mentioned in chapter 1, the allomorph [kfʲatɕ] does occur in a derived place name *kwiaciarnia* ‘florist’s shop’, as well as *kwiaciarka* ‘(lady) florist’, which recently replaced the original *kwieciarnia* (*kwieciarka*). A possible explanation lies in the fact that from the semantic point of view, ‘florist’s shop’ and ‘florist’ are in correspondence with ‘flower’ (phonologically associated with its major allomorph [kfʲat]) and not with derivatives mentioned earlier, as ‘flowery’, ‘flower bed’ etc., with which it shares the phonological “soft/front” environment. The analogical replacement of the stem vowel makes the semantic relation more transparent and makes the words *kwiaciarnia* and *kwiaciarka* bluntly obvious.

Let us now turn to the analogically leveled nouns from table IX. Many of them lack derivatives based on the *e*-variant of the stem (and have either no derivatives at all or have ones with the back vowel suffix), namely: *lód* ‘ice’, *dziad* ‘pauper’, *zwierciadło* ‘mirror’, *siodło* ‘saddle’, *wiadro* ‘bucket’, *wiosło* ‘paddle’ and *wrzód* ‘ulcer’. Some others have derivatives of low frequency in which the vowel is historically documented (or reconstructed with high probability) as *e* and which was analogically leveled to *a/o*, cf. *siostrzeniec* ‘nephew’, *przedwiośnie* ‘early spring’, *miodny* ‘melliferous’, *miotlasty* ‘broom-shaped’, *biesiadnik* (*biesiadny*-adj.) ‘reveller’. In one case, an *e*-derivative is of rare occurrence and of some semantic distance too, namely *brzezina* ‘birch wood’. In another case, a semantically closer derivative is analogical and a more distant derivative preserves the *e*-stem, cf. *jeziorny* ‘lake-adj.’, *pojezierze* ‘lake district’. As to the

<sup>12</sup> This is *contra* the claims on analogy made in the OT literature, cf. chapter 8.

<sup>13</sup> In the OT framework, this effect can be attributed to an unviolable constraint, which enforces minimal faithfulness violations. Polish consonants often alternate between their „hard” and „soft” versions in a number of contexts, depending on the palatalizing/non-palatalizing property of the suffix.

remaining nouns, four have productive *e*-stem derivatives of high occurrence but independent meanings, loosely connected with the meanings of the base nouns, cf. *czas* ‘time’ – *wczesny* ‘early’, *współczesny* ‘contemporary’, *doczesny* ‘earthly’; *żona* ‘wife’ – *żeński* ‘female’, *żenić się* ‘get married’, *małżeństwo/ożenek* ‘marriage’; *oddział* ‘department’ – *oddzielny* ‘separate-adj.’, *oddzielać* ‘to separate’; *śląd* ‘trace’ – *śledzić* ‘follow, stalk’. In the entire group, there is one noun *ściana* ‘wall’, which has a semantically transparent *e*-stem derivative *ścienny* ‘wall-adj.’ With this one exception, the nouns which have undergone leveling do not have derivatives that could support the minor *e*-variant.

The above examples of derivatives indicate that the closeness/distance in meaning positively/negatively correlates with stem leveling. The same principle is observable in the case of lexical splits which affected the locatives of three nouns listed in table VIII. In each case, an *e*-variant is used in a fixed idiomatic expression and an analogical *a/o*-variant in the predictable, regular meaning of the noun. Thus, the locative *czole* is used in the meaning of ‘forehead’, as in *masz coś na czole* ‘you’ve got something on your forehead’, while *czele* is used in the abstract sense of ‘head’, as in *iść na czele pochodu* ‘lead a parade’ (lit. ‘walk at the forehead of a parade’), *stać na czele partii* ‘lead a (political) party’ (lit. ‘stand at the forehead of a party’). In the case of the next noun, the locative *przodzie* is used in the meaning of ‘front part (of something)’, e.g. *na przodzie sukienki (łodzi, autobusu)* ‘in front (part) of a dress (boat, bus)’ and *przedzie* is used without a nominal complement in a more abstract adverbial sense of ‘in front’, e.g. *iść na przodzie* ‘walk in front’. It is worthwhile to point out that the analogical form *przodzie* may be occasionally used instead of *przedzie*, unlike in the previous case of *czole* and *czele* which cannot be used interchangeably. The reason for this difference seems rather obvious: the two ‘fronts’ are much closer semantically than ‘forehead’ and ‘leadership’. The last of the nouns, *miara*, may have a concrete meaning of ‘measure’ as an ‘instrument for measuring’, as well as an abstract meaning of ‘degree’. The latter occurs in numerous and frequent expressions (1369 times in the PWN Corpus), as *w dużej (znacznej, pewnej, równej, żadnej etc.) mierze* ‘in great (considerable, some, equal, any) measure’. In the corpus, there was also one occurrence of *miarze* in the “concrete” meaning in the following context: *w miarze wypitego alkoholu* ‘in the measure (i.e. quantity) of the drunk alcohol’.

The last piece of data to be discussed is the few nouns in which analogy took the opposite direction, i.e. towards the minor *e*-form. I have identified two such cases, as well as two others, which were a result of lexical splits.

It is historically documented that the nouns *cena* ‘price’ and *krzesło* ‘chair’ had alternating paradigms still in the 15<sup>th</sup> century, with nominative *\*cana* and *\*krzasło*, respectively (cf. Bańkowski 2000, Brückner 1974). At a later time, the *a*-stem vowel of nominative and other cases was replaced by the *e* occurring only in the singular locative (locative=dative of *cena*)<sup>14</sup>. According to the PWN Corpus, the two nouns have the following frequencies of their WORD and locative forms: CENA 30552, cenie 1858 (6,08%) and KRZESŁO 1128, krześle 238 (21,10%). As can be seen, the frequencies of the (historical) *e*-forms are significant enough to *prevent* the analogy towards the (historically) major *a*-pattern, but are not high enough to *trigger* the analogy towards the

<sup>14</sup> It is interesting to observe how analogy caused these words to „go back” to their proto-forms, cf. early Pol. *\*tsena* (before the Vowel Shift) > *\*tsana* (after Vowel Shift) > *tsena* (after analogy).

*e*-pattern. However, it is very likely that, at least in the case of *cena*, the occurrence of the *e*-stem variant was higher than the occurrence of the *a*-variant at the time the analogy actually took place (ca. 16<sup>th</sup> c.). The most common uses of this word are obviously connected to asking about the price of something. In contemporary Polish, the most typical expressions used for this purpose occur either with the nominative *cena*, the accusative *cenę* or the locative *cenie*, cf. *jaka jest tego cena*, or *jaką to ma cenę*, or *jakiej to jest cenie* ‘what’s the price of this’. Similarly, the expressions for telling the price vary and may include different declensional cases, out of which the locative is a possibility, but not the most frequent one. This is the reason why the locative word-form has a relatively small percentage of occurrence in the PWN Corpus. However, the data of 16<sup>th</sup> c. Polish suggest that expressions with the locative case *być w cenie* ‘be in price (of)’ and *być w jakiej cenie* ‘be in what price’ were the most common way of telling and asking about the price (cf. *Słownik* 1968-2004). The same expression *być w cenie* was also commonly used in a metaphorical meaning ‘have a value’, as well as *mieć w cenie* lit. ‘have in value’ i.e. ‘respect’. Taking these facts into account, we can hypothesize that at that time the use of the *e*-variant could be higher than the use of the *a*-variant. In addition, the (historical) *e*-stem occurs in commonly used derivatives (frequent in Old Polish, too) as the verb *cenić* ‘value’ and the adjective *cenny* ‘valuable’ (and a few others), which provided additional support for this direction of analogy, especially in the absence of the *a*-stem derivatives.

As to the word *krzesło* ‘chair’, it does not have any derivatives and it has not undergone any significant change in meaning and usage which could shed light on the direction of analogy towards the *e*-form. Since phrases associated with sitting *na krześle* ‘on the chair’ are common in everyday use, it is possible that the percentage of the locative would turn out much higher if the data of the spoken corpus were considered. If so, the high frequency could explain why the locative served as a base for analogy. However, in the absence of such data, no sensible explanation can be offered. According to historical sources, the analogical replacement took place in 16<sup>th</sup>/17<sup>th</sup> c., but a small number of dialects preserved the older *krzasło* (Bańkowski 2000 vol. 1:835).

Lexical splits took place in two nouns, namely, *biada* – the exclamation of misfortune (rarely used in the contemporary language) versus *bieda* ‘poverty’, and *działo* ‘cannon’ versus *dzielo* ‘work (composition)’. It is hard to explain for certain why the semantic splits which affected these particular words were accompanied by phonological splits. Nevertheless, given that the splits *did* take place, we can understand why they happened in the way they did and not vice versa.

As to the former case, the dialectal innovative form *bieda* started to spread onto other dialects around the 17<sup>th</sup> century, when the word was already polysemous: it covered the original meaning of ‘misfortune’, as well as the newer sense of ‘poverty’. The adjective *biedny* was in common use as ‘poor’ and the obsolete noun *biednik* ‘poor person’ (cf. Sławski 1952-1974, vol. 1: 31-32). This created strong support for the leveling of the *e*-form in the meaning of ‘poor’ (cf. Bańkowski 2000, vol. 1: 46). The exclamation *biada* ‘misfortune’ remained immune to the new replacement (presumably, also in the dialects which initiated the change), because of a distinct meaning. In sum, the split here is very much alike the splits *czele~czole* etc. discussed previously.

As to the *działo~dzielo* example, the facts are less clear. Bańkowski suggests that the split in Polish (ca. 15<sup>th</sup> c.) could have been influenced by a similar development that

took place in Czech (cf. Bańkowski 2000, vol. 1: 325). In any event, the original *e*-form was found only in the locative of the originally polysemous word meaning ‘cannon’ and ‘work (composition)’. While the locative expressions, such as *w (tym) dziele* ‘in (this) work’ or *o (tym) dziele* ‘about (this) work’ are (and were in Old Polish) common while talking about the contents of a given work (of art, music, literature etc.), it is hard to imagine a natural context of use of the locative in the sense of ‘canon’.<sup>15</sup> All usual contexts, associated with shooting, loading or cleaning the weapon, would employ other declensional cases, cf. *strzelać z działa* (gen.) ‘shoot from the canon’, *ładować (czyścić) działo* (acc.) ‘load (clean) the cannon’. We can presume that at the time of the split, the locative was more often used in the meaning of ‘work’ than ‘cannon’, hence the *e*-form was more salient for the former and not for the latter.

#### 2.4. An Optimality Theoretic analysis of the data

The model of extended correspondence constraints can account for the finest distinctions among particular words as to their behavior towards the *e~a/o* alternation. Likewise, it can capture similarities by grouping various items into lexical “clusters”. An OT analysis consists in appropriate ranking of the correspondence constraints pertaining to these clusters with respect to an “alternation” constraint. I assume that the latter can be defined in morphophonological terms as prohibiting sequences such as *C’oT’* and *C’aT’* (where *C’* denotes a “soft” consonant and *T’* a “soft” coronal) before particular suffixes, including the inflectional suffix *-e* of the locative singular and several derivational suffixes, e.g. the adjectival suffix *-n-* (historically preceded by the soft “yer”). This rough formulation (NB reversing the historical change) is more accurate than an attempt at stipulating purely phonological conditioning, which would be often quite untenable, as, for example, in the case of the above mentioned suffix *-n-*. Alternatively, the *E~A/O* constraint can be stated in templatic terms, requiring a “matching” sequence of *C’eT’* in specific environments such as the locative singular, adjectives suffixed with *-n-*, etc. This formulation of the constraint can also account for few words with a three-part paradigmatic alternation such as in *kości[u]ł* ‘church-nom.’, *kości[ɔ]ła* ‘church-gen.’, *kości[e]le* ‘church-loc.’ (cf. chapter 3). In any case, the *E~A/O* constraint must be ranked relatively low, i.e. just below *Cor* to guarantee minimal violation of correspondence. I will illustrate the interaction of the system using examples of several stems discussed earlier in this chapter, including in the analysis correspondence constraints operating not only among intraparadigmatic members, but also among base nouns and their derivatives.

Let us first look at the leveled nouns, using the examples of *jezioro* ‘lake’, *czas* ‘time’ and *żona* ‘wife’ in (11) below. Their intraparadigmatic correspondence constraints are included in clusters *Cor-A<sub>1</sub>*, *Cor-A<sub>2</sub>* and *Cor-A<sub>3</sub>*, respectively. They must dominate *E~A/O* since the analogous locatives as *jeziorze*, *czasie*, *żonie* ignore the morphophonological requirement. Correspondence constraints involving derivatives differ in each group. In the case of the noun ‘lake’, there exists a rare adjective *jeziorny* with its vowel leveled by analogy. This is captured by high ranking of the relevant cluster *Cor-B<sub>1</sub>* in (11a). In the case of the semantically more distant derivative, *pojezierze* ‘lake district’, there is no correspondence with the base noun (or derived adjective), hence the

<sup>15</sup> Some such attempts include: *(mówić) o dziale* ‘(talk) about the cannon’, *(stać) na dziale* ‘(stand) on the cannon’, *(utkwić) w dziale* ‘(be stuck) in the cannon’ (although in the last example a more precise location would be probably needed for communicative purposes).

placement of Cor-C<sub>1</sub> below E~A/O. The remaining two nouns have more derivatives and, consequently, more finely grained constraint clusters. Some derivatives are semantically close, but have their stem vowel “incidentally” identical to that of the base nouns, since they are formed with back suffixes. Because they do not interact with the E~A/O constraint, their correspondence constraints may have either ranking, as shown by parentheses in (11b) and (11c). They include the adjective *czasowy* ‘temporary’ and the compound *czasomierz* ‘timer’ in Cor-B<sub>2</sub>, as well as diminutives such as *żonka*, *żoneczka*, *żonusia* ‘wife-dim.’ or the participle *żonaty* ‘married (man)’ in Cor-B<sub>3</sub>. The derivative included in Cor-C<sub>2</sub>, *wczesny* ‘early’, is semantically distant (quite unpredictable) from the base noun which goes together with the lack of phonological correspondence. However, the verb in Cor-C<sub>3</sub>, *żenić* (*się*) ‘to marry’ as well as some further derivatives, as e.g. *bezzenny* (*stan*) ‘(state of) not being married’, are semantically transparent. The fact that they do not correspond with the base noun (but they do correspond with each other) can be explained by the high frequency of the verb, which prevents analogy. The final clusters in (11b) and (11c) include all kinds of semantically loose relations among stem-sharing words with either identical or different stem vowels. Because of the lack of common semantics, it seems appropriate to place these Cor constraints low in the hierarchy, even if they “incidentally” share the stem vowel. The examples include: *wczasy* ‘package vacation’ and *czasownik* ‘verb’ in (11b) as well as *żeński* ‘female’ in (11c).

(11) Correspondence constraints of *jezioro* ‘lake’, *czas* ‘time’ and *żona* ‘wife’

a/ Ranking: Cor-A<sub>1</sub> 4 Cor-B<sub>1</sub> >> E~A/O >> Cor-C<sub>1</sub>

Cor-A<sub>1</sub> {*jeziorze: jezioro, jeziorze: jeziora, jezioro: jeziora* etc.}

Cor-B<sub>1</sub> {*jeziorny: jezioro, jeziorny: jeziora* etc.}

Cor-C<sub>1</sub> {*pojezierze: jezioro, pojezierze: jeziorny, jezioro: pojezierze* etc.}

b/ Ranking: Cor-A<sub>2</sub> (Cor-B<sub>2</sub>) >> E~A/O >> Cor-C<sub>2</sub> 4 Cor-D<sub>2</sub> (Cor-B<sub>2</sub>)

Cor-A<sub>2</sub> {*czasie: czas, czasie: czasu, czas: czasu* etc.}

Cor-B<sub>2</sub> {*czasowy: czas, czas: czasowy, czasomierz: czas* etc.}

Cor-C<sub>2</sub> {*wczesny: czas, czas: wczesny* etc.}

Cor-D<sub>2</sub> {*wczasy: czas, wczasy: wczesny, czasownik: czas* etc.}

c/ Ranking: Cor-A<sub>3</sub> (Cor-B<sub>3</sub>) >> E~A/O >> Cor-C<sub>3</sub> 4 Cor-D<sub>3</sub> (Cor-B<sub>3</sub>)

Cor-A<sub>3</sub> {*żonie: żona, żonie: żony, żona: żony* etc.}

Cor-B<sub>3</sub> {*żonka: żona, żona: żonka, żonaty: żona* etc.}

Cor-C<sub>3</sub> {*żenić się: żona, żona: żenić się, bezzenny: żona* etc.}

Cor-D<sub>3</sub> {*żeński: żona, żeński: żenić się* etc.}

The next set of examples listed in (12), contains alternating base nouns and their derivatives. A simple case is that of *las* ‘forest’ in (12a), since all correspondence constraints follow the E~A/O morphophonology. Intraparadigmatic constraints are

included in the Cor-A<sub>i</sub> cluster, while Cor-B<sub>i</sub> includes adjectives as *leśny* ‘forest’, *lesisty* ‘forested’, nouns as *leśnik* ‘forester’, *leśniczówka* ‘forester’s lodge’, etc. Similarly in (12b): intraparadigmatic constraints of *kwiat* ‘flower’ (Cor-A<sub>j</sub>), as well as those pertaining to derivatives, e.g. *kwiecisty* ‘flowery’, *ukwiecić* ‘decorate with flowers’, *kwietnik* ‘flower bed’ (Cor-B<sub>j</sub>), are dominated by E~A/O. However, as discussed in the previous section of this chapter, the situation of this stem is somewhat exceptional, since the locative *kwiecie* has unusually low frequency in the absolute sense and in the relative sense (i.e. with respect to the major allomorph), and is then much less salient than the minor locative case of other nouns of this small alternating class. We can expect then that analogy may soon take place and Cor-A<sub>j</sub> will move above E~A/O. Such reranking has already happened with respect to two nouns semantically closer to *kwiat* ‘flower’ than other derivatives, namely *kwiaciarnia* ‘florist’s shop’ and *kwiaciarka* ‘(woman) florist’, as discussed in the previous section. I distinguish these strong correspondence constraints in (12b) as Cor-B’<sub>j</sub>, which include the major allomorph of ‘flower’ in the Base position of the correspondence relation.

(12) Correspondence constraints of *las* ‘forest’ and *kwiat* ‘flower’

a/ Ranking: E~A/O >> Cor-A<sub>i</sub> 4 Cor-B<sub>i</sub>

Cor-A<sub>i</sub>{*lesie: las, las: lesie, las: lasu* etc.}

Cor-B<sub>i</sub>{*leśny: las, lesisty: las, leśnik: las* etc.}

b/ Ranking: Cor-B’<sub>j</sub> >> E~A/O >> Cor-A<sub>j</sub> 4 Cor-B<sub>j</sub>

Cor-A<sub>j</sub>{*kwiecie: kwiat, kwiat: kwiecie, kwiat: kwiatu* etc.}

Cor-B<sub>j</sub>{*kwiecisty: kwiat, ukwiecić: kwiat, kwietnik: kwiat* etc.}

Cor-B’<sub>j</sub>{*kwiaciarnia: kwiat, kwiaciarnia: kwiatu, kwiaciarka: kwiat* etc.}

It should be noted that all the constraint clusters presented above do not have any formal status, but simply facilitate the discussion. In the overall grammar, such clusters, which refer to particular stems, create larger units – lexical classes, such as a non-alternating class {A<sub>1</sub> 4 A<sub>2</sub> 4 A<sub>3</sub>...etc.} and an alternating class {A<sub>i</sub> 4 A<sub>j</sub> 4 A<sub>k</sub>...etc.}.

## CHAPTER 3

# The *o~u* alternation in Polish nouns

### 3.1. The historical source of the alternation and its current scope

Polish originally had nine pairs of short/long vowels and two extra-short “yers”. The opposition between short and long vowels was maintained until the 15<sup>th</sup> century. In the early 16<sup>th</sup> c. long vowels shortened but the process was accompanied by raising of the non-high vowels. As Rospond (2003:48) points out, in the 16<sup>th</sup> c. poetry the reflex of the old \**ō* still rhymed with [o], but in the 17<sup>th</sup> c. it started to rhyme with [u]. The slight quality distinction between [u] and the reflex of the historical \**ō* continued throughout the next centuries until they finally merged in the 19<sup>th</sup> c. (Długosz-Kurczabowa and Dubisz 1998:129). Mergers also took place in the case of two other historically long vowels (but until now there exist dialects which preserve the oppositions in one way or another). Polish orthography is conservative in marking the historical \**ō* as *ó* [u], while the [u] resulting from \**u* or \**ū* is spelt as *u*.

One of the sources of long vowels in Old Polish was compensatory lengthening (with a functional touch) which accompanied the process of yers’ deletion (cf. chapter 4). The loss of yers in the word-final position created closed syllables in which phonemically short vowels were subject to a typical process of non-phonemic lengthening before a voiced coda. The emergence of the long/short opposition in the contrasting environment of a voiced/voiceless coda compensated for the weakened voice feature of the final consonant, which was presumably already affected by final devoicing (cf. Furdal 1964:54). The vowel lengthening increased the contrast between certain lexical pairs, as illustrated in (13) below, but at the same time it caused the long/short alternation of the stem vowel, since before a full vowel suffix (i.e. in an open syllable) the lengthening did not occur. After the raising of the long *ō* to *u*, the Old Polish *o~ō* alternation became *o~u* alternation and is still found in stems ending in a phonemic voiced consonant (except for nasals<sup>16</sup>).

#### (13) The source of the *o~u* alternation

Early Polish	Old Polish	modern Polish	gloss
a/ rokъ, roġъ	rok, rōġ (rōk)	rok, ruk	‘year’, ‘horn’ (nom.)
roku, roġu	roku, roġu	roku, roġu	‘year’, ‘horn’ (gen.)
b/ bokъ, boġъ	bok, bōġ (bōk)	bok, buk	‘side’, ‘god’ (nom.)
boku, boga	boku, boga	boku, boga	‘side’, ‘god’ (gen.)

<sup>16</sup> According to Długosz-Kurczabowa and Dubisz (1998: 101n.), the lengthening took place before nasals too, but then *u* was replaced by *o* in the standard language as a result of hypercorrection. Exceptionally, the alternation is found before a phonemically voiceless consonant, cf. *stopa* ‘foot-nom.’, *stóp* ‘feet-gen.’ and three other stems, cf. Krzyżanowski (1992:54). The fixed *u* or *o* before a voiced consonant typically appears in loan-words and very few native words (see Tokarski 2001:62n. for examples), sometimes resulting from analogy, as shown later in this section.

Once the length contrast disappeared and was replaced by quality changes, the alternation lost its phonetic motivation and transparency and its role as a productive rule diminished. But the alternations remained with great regularity in a large number of nouns, as well as in other categories (cf. *robić* ‘to do’, *rób* ‘do-imp.’). The paradigms below contain examples of nouns of all three genders, with the masculine gender further divided into “inanimate” (represented by *wóz* ‘cart’) and “animate” (represented by *bóg* ‘god’) declension types. As we can see, the *u*-allomorph (underlined) appears as a minor pattern, especially for feminine and neuter nouns in which it occurs only in one declensional case of low frequency, namely the genitive plural. In masculine nouns, the *u*-allomorph is slightly more salient since it occurs in the frequent form of the nominative singular (of all nouns), as well as in the accusative (which is equal to the nominative) of the inanimate type.

(14a) The singular paradigm of *o~u* nouns

case	masculine	feminine	neuter
nom sg	<u>wóz</u> , <u>bóg</u>	głowa	słowo
gen sg	wozu, boga	głowy	słowa
dat sg	wozowi, bogu	głowie	słowu
acc sg	<u>wóz</u> , boga	głowę	słowo
instr sg	wozem, bogiem	głową	słowem
loc sg	wozie, bogu	głowie	słowie
voc sg	wozie, boże ‘cart’, ‘god’	głowo ‘head’	słowo ‘word’

(14b) The plural paradigm of *o~u* nouns

case	masculine	feminine	neuter
nom pl	wozy, bogowie	głowy	słowa
gen pl	wozów, bogów	<u>głów</u>	<u>słów</u>
dat pl	wozom, bogom	głowom	słowom
acc pl	wozy, bogów	głowy	słowa
instr pl	wozami, bogami	głowami	słowami
loc pl	wozach, bogach	głowach	słowach
voc pl	wozy, bogowie ‘carts’, ‘gods’	głowy ‘heads’	słowa ‘words’

It has been shown previously in chapter 2 that a minor allomorph was very much susceptible to analogy, especially for low frequency WORDs and word-forms. No such effect can be observed in the case of the *o~u* alternating nouns. Whether their frequency is high, medium or low, they show a stable alternating pattern. Stem leveling does not take place even in nouns of very low frequency, such as e.g. the following (with the PWN Corpus figures for WORD and the minor word-form indicated in brackets); masculine gender: *GLÓG* (87) & *glóg* (5) ‘hawthorn’, *BARŁÓG* (45) & *barłóg* (13) ‘lair; shabby

bed’, *ZNÓJ* (76) & *znój* (33) ‘toil’; and feminine gender: *PŁOZA* (28) & *plóz* (0) ‘runner (of sleigh)’, *MORGA* (38) & *mórg* (10) ‘unit of area (ca. 1,4 acre)’. Very few native words underwent leveling and have alternating forms attested only historically, e.g. *jęzor* ‘(big) tongue’ (\**jęzór*), *muchomor* ‘fly agaric’ (\**muchomór*), *jezior* ‘lakes-gen.’ (\**jeziór*). Likewise, pattern-analogy very rarely affects borrowings, cf. *moda* ‘fashion-nom.’, *mód* ‘fashions-pl.’, or *pagoda* ‘pagoda-nom.’, *pagód* ‘pagodas-gen.’. Normally, loan-words do not exhibit alternation, as can be seen on the example of the following nouns, which very much resemble the two previous ones in their phonological make-up: *metoda* ‘method-nom.’, *metod* ‘methods-gen.’ and *anoda* ‘anode-nom.’, *anod* ‘anodes-gen.’.

To conclude, the *o~u* alternation characterizes a certain portion of the lexicon and does not show a tendency either to narrow down – by means of stem analogy affecting less frequent words, or to extend – by pattern analogy applicable to new lexical items. The former fact can be attributed to a relatively large size of the lexical class exhibiting the alternation, which creates a strong, salient morphophonemic pattern. Since it is easily retrievable, stem analogy does not have to take place. Why is not the alternating pattern mapped onto other lexical items, for example, recent loan-words? The reason here is perhaps far from obvious, but it seems that not much would be gained if the pattern did extend. In the particular case here, the burden of stem-alternation would not be compensated for by a decrease in markedness. In OT terms, there is no higher constraint that would impose O-O stem-correspondence violation.

The following section concentrates on the distribution of the alternating nouns in the lexicon. As in the previous chapter, the frequency dictionary (*Słownik* 1990) will be used as a source.

### 3.2. The distribution of alternating nouns in the lexicon

Table X below comprises 33 alternating nouns spotted in the first thousand of the most frequent words. For comparison, table XI shows nouns containing non-alternating stem *o* or *u* (orthographic *u* or *ó*) occurring in the same environment of voiced consonant (excluding nasals). Out of the three words found within the same range, two are loan-words and one is a diminutive, subject to analogical stem leveling (cf. later in this chapter). We can see that in the most frequent vocabulary, the number of alternating nouns is significantly higher than the number of non-alternating ones.

Table X. The *o~u* alternating nouns (with the gender indicated), ordered according to their rank, as they appear in the first 1000-word list.

word (nom.) + gloss	rank	frequency
sposób (m) ‘manner’	142	351
szkoła (f) ‘school’	150	334
woda (f) ‘water’	155	328
rozwój (m) ‘development’	158-159	320
głowa (f) ‘head’	204	255
słowo (n) ‘word’	205	254

budowa (f) 'structure'	206	250
naród (m) 'nation'	269-270	205
osoba (f) 'person'	279-281	196
pokój (m) 'room'	282-284	195
rola (f) 'role'	355-358	162
zespół (m) 'team'	372	154
rozmowa (f) 'conversation'	375-378	152
bóg (m) 'god'	379-381	151
samochód (m) 'car'	412-416	143
wieczór (m) 'evening'	455-459	132
noga (f) 'leg'	474-476	126
koło (n) 'wheel'	491-496	122
załoga (f) 'crew'	565-578	109
choroba (f) 'disease'	579-583	108
pole (n) 'field'	579-583	108
obóz (m) 'camp'	632-638	99
powód (m) 'reason'	639-649	98
wybór (m) 'choice'	650-651	97
czoło (n) 'forehead'	736-749	86
morze (n) 'sea'	765-778	83
zbiór (m) 'collection'	779-787	82
wóz (m) 'cart'	846-854	76
wyrób (m) 'product'	878-892	73
zawód (m) 'profession'	904-913	71
spokój (m) 'calmness'	940-958	68
dowód (m) 'proof'	982-1001	66
stół (m) 'table'	982-1001	66

Table XI. Nouns with non-alternating *o*, *ó* [u] or *u* in the first 1000-word list.

word (nom.) + gloss	rank	frequency
metoda (f) 'method' (loan)	432-436	138
próba (f) 'attempt' (loan)	565-578	109
kółko (n) 'circle (dim.)'	846-854	76

The following tables XII, and XIII, present respective lists of nouns in the second thousand of words in *Słownik* (1990). They contain 25 alternating nouns (including two before a voiceless consonant) and 14 non-alternating ones.

Table XII. The *o~u* alternating nouns ranked between 1003-2009.

word (nom.) + gloss	rank	frequency
obrót (m) ‘turn’	1047-1065	62
środa (f) ‘Wednesday’	1047-1065	62
dół (m) ‘ditch’	1066-1077	61
mowa (f) ‘speech’	1128-1147	57
wzór (m) ‘pattern’	1128-1147	57
dochód (m) ‘income’	1186-1209	54
pora (f) ‘time’	1244-1270	51
krowa (f) ‘cow’	1271-1301	50
zboże (n) ‘cereal’	1322-1355	48
pogoda (f) ‘weather’	1356-1379	47
wschód (m) ‘East’	1356-1379	47
stopa (f) ‘foot’	1380-1413	46
opór (m) ‘resistance’	1444-1475	44
wola* <sup>17</sup> (f) ‘will’	1476-1505	43
powrót (m) ‘return’	1506-1540	42
umowa (f) ‘agreement’	1506-1540	42
lód (m) ‘ice’	1541-1570	41
zgoda* (f) ‘agreement’	1541-1570	41
ustrój (m) ‘political system’	1571-1612	40
dobro (n) ‘right’, pl. ‘property’	1613-1656	39
ogród (m) ‘garden’	1613-1656	39
przyroda* (f) ‘nature’	1657-1697	38
zasób (m) ‘supply’	1657-1697	38
utwór (m) ‘work, composition’	1698-1757	37
podłoga (f) ‘floor’	1758-1805	36
szkoda (f) ‘damage’	1758-1805	36
zachód (m) ‘West’	1758-1805	36
nawóz (m) ‘fertilizer’	1910-1941	33
prośba (f) ‘request’	1910-1941	33
sól (f) ‘salt’	1942-2008	32

<sup>17</sup> The feminine nouns marked with a star do not occur in testable forms (diminutives or plural gen.) to decide for certain, whether they alternate or not. I have included them for the following reasons: *wola* is used as a common place-name which also occurs in a diminutive form as *Wólka*; *zgoda* shows alternation in a fairy-tale name *Niezgódka*; *przyroda* contains an easily identifiable alternating root *rod~ród* (hence, if someone wished, “*Przyródka*” could make a good name for another fairy).

Table XIII. Nouns with non-alternating *o*, *ó* [u] or *u* ranked between 1003-2009.

struktura (f) ‘structure’ (loan)	1302-1321	49
skóra (f) ‘skin’	1322-1355	48
ambasador (m) ‘ambassador’ (loan)	1356-1379	47
natura (f) ‘nature’ (loan)	1380-1413	46
podróż (f) ‘journey’	1380-1413	46
lud (m) ‘people’	1414-1443	45
ból (m) ‘pain’	1506-1540	42
reguła (f) ‘rule’ (loan)	1506-1540	42
trud (m) ‘hardship’	1506-1540	42
wagon (m) ‘carriage’ (loan)	1506-1540	42
wódka (f) ‘vodka’ (dim.)	1571-1612	40
jezioro (n) ‘lake’	1806-1853	35
kula (f) ‘ball’	1910-1941	33
mur (m) ‘wall’ (loan)	1910-1941	33

The disproportion between the class of nouns with the alternating vowel and those with the fixed *o*, *ó* [u] or *u* (before the “voiced” consonant) continues throughout medium frequencies but it slowly diminishes once we reach low frequencies. Within the range of four-occurrence words in *Słownik* 1990 (ranks from 8739-10355), I have counted 14 alternating nouns (all native, e.g. *ozdoba* ‘ornament’, *potwór* ‘monster’, *topola* ‘poplar’, *wrzód* ‘ulcer’) and 7 non-alternating (native or borrowed, e.g. *pomidor* ‘tomato’, *senior* ‘senior’, *tura* ‘round’, *śluga* ‘servant’, *zguba* ‘loss’).

The distributional data indicate that the type frequency of the *o~u* alternating nouns is significant enough to create a strong pattern, which resists stem leveling. In fact, the alternation constitutes the dominant pattern (in the “voiced” consonant environment) given the relative rarity of the non-alternating nouns, especially in higher frequencies. In order to visualize the size and the frequencies of the *o~u* nouns, let us briefly compare them to the previously discussed *e~a/o* nouns. Table XIV presents the number of occurrences of the nouns of these two classes for various frequencies (the so-called “alternable” nouns are historically alternating, i.e.: presently alternating and analogically leveled, listed in this order in table XIV).

Table XIV. The occurrences of the *o~u* alternating and *e~a/o* “alternable” nouns.

	<i>o~u</i> alternating	<i>e~a/o</i> “alternable”
ranks 1-1002 in <i>Słownik</i> 1990	33	16 (10+6)
ranks 1003-2009 in <i>Słownik</i> 1990	25	13 (5+8)
ranks 8739-10355 in <i>Słownik</i> 1990	14	1 (0+1 <i>wiadro</i> ‘bucket’)

As we can see from the above table, the *o~u* alternating nouns are significantly more frequent than the *e~a/o* “alternable” nouns in all frequency ranks. This fact should not come as a surprise if we take into consideration phonological environments of these

alternations. The context for the *o~u* alternation is much less restrictive, namely: before any voiced consonant (except for nasals); the environment for the *e~a/o* alternation is extremely narrow: after a “soft” consonant and before a coronal (further restricted to alternating between a palatalized and a non-palatalized variant). Naturally, the number of lexical stems which fulfill the first condition is larger than the number of those that fulfill the second one. Consequently, the difference in productivity between the two alternations directly relates to the difference in size between the relevant lexical classes, but indirectly relates to the phonological process itself – its general or more specific character.

It is worthwhile to observe that the environments of these two alternations overlap in the case of stems having a palatalized consonant followed by *o*, followed in turn by a final voiced consonant. If a noun of that particular shape belongs to the infrequent vocabulary, its stem vowel alternates between *o~u*, but not between *e~o*, cf., among others (with the analogical form underlined), *miód*, *miodu*, *miodzie* ‘honey-nom., gen., loc.’, *lód*, *lodu*, *lodzie* ‘ice-nom., gen., loc.’ *wróż*, *wrządu*, *wrzodzie* ‘ulcer-nom., gen., loc.’, *brzoza*, *brzozie*, *brzóz* ‘birch-nom., loc. gen. pl.’ This effect does not need to be specially stipulated, but simply follows from the frequency criteria and the analogy threshold of the unproductive pattern.<sup>18</sup>

Let us now turn to the part of the lexicon, in which analogy apparently does take place with respect to the *o~u* alternation.

### 3.3. The problem of diminutives

Diminutives in Polish can be formed with a number of suffixes out of which the suffix *(e)k* appears as the most productive. Derivatives with this particular suffix are relevant to our discussion here, because, due to the suffix’ alternation between *ek* and *k*, the stem vowel occurs either in an open or in a closed syllable within the paradigm. Unlike in the case of unsuffixed base nouns, however, the stem vowel in diminutives does not alternate between *o~u*, but is fixed throughout the declensional paradigm, as illustrated in (15). In feminine and neuter nouns, the vowel is always *u*, even though in the genitive plural it occurs in an open syllable, where *o* is expected. In the masculine paradigm, the stem vowel is usually *o*, but for some lexically specified words it is *u*; nevertheless, it does not alternate within the paradigm, either. Depending on which vowel is leveled in the masculine paradigm, each particular word-form appears as “regular” or “irregular” with respect to the *o~u* alternation. I will argue that all “irregular” word-forms (underlined in the paradigms below) are analogical.<sup>19</sup>

(15a) The singular paradigm of the diminutives of *o~u* nouns

<sup>18</sup> Naturally, high frequency nouns of the required phonological shape are predicted to alternate with respect to both rules. There seem to be two such nouns, namely, *kościół*, *kościola*, *kościelie* ‘church-nom., gen., loc.’ and *popiół*, *popiołu*, *popiele* ‘ash-nom., gen., loc.’ (but the latter is not a frequent word anymore). In one stem the vowel *o* is found in the nominative instead of the expected *u*, cf. *anioł*, *aniola*, *aniele* ‘angel-nom., gen., loc.’ (likewise *archanioł* ‘archangel’). I can not think of any explanation of this behavior, but it should be noted that this early loan-word had a rather peculiar development (e.g. according to some sources, the stem vowel was still *o* in the 14th c., cf. Bańkowski 2000 v.1:12).

<sup>19</sup> The analogy-based analysis of Polish diminutives was earlier proposed in Kraska-Szlenk (2003/1995) and (1999a), cf. also Benua (1997), Kenstowicz (1996). Phonological attempts at explaining these data are very problematic, as discussed in Kraska-Szlenk (2003:59-60).

case	masculine	feminine	neuter
nom sg	<u>wózek</u> , bożek	główka	słówko
gen sg	wózka, <u>bożka</u>	główki	słówka
dat sg	wózkowi, <u>bożkowi</u>	główce	słówku
acc sg	<u>wózek</u> , <u>bożka</u>	główkę	słówko
instr sg	wózkiem, <u>bożkiem</u>	główką	słówkiem
loc sg	wózku, <u>bożku</u>	główce	słówku
voc sg	wózku, <u>bożku</u>	główko	słówko
	‘cart’, ‘god’	‘head’	‘word’

(15b) The plural paradigm of the diminutives of *o~u* nouns

case	masculine	feminine	neuter
nom pl	wózki, <u>bożki</u>	główki	słówka
gen pl	wózków, <u>bożków</u>	<u>główek</u>	<u>słówek</u>
dat pl	wózkom, <u>bożkom</u>	główkom	słówkom
acc pl	wózki, <u>bożki</u>	główki	słówka
instr pl	wózkami, <u>bożkami</u>	główkami	słówkami
loc pl	wózkach, <u>bożkach</u>	główkach	słówkach
voc pl	wózki, <u>bożki</u>	główki	słówka
	‘carts’, ‘idols’	‘heads’	‘words’

Diminutive derivation in Polish is very productive. Practically, all nouns denoting concrete objects, persons, animals, plants, etc. can occur in some form of diminutive, and those that cannot are mostly abstract or collective nouns. In order to have a rough idea of the productivity of the diminutive formation, let us consider a sample of 56 nouns listed earlier in tables I and II, included between ranks 1 and 2009 in *Słownik* (1990). Almost half of these nouns (25) form diminutives with the *(e)k* suffix<sup>20</sup>, for example, those listed in (15) above (cf. also table V below). Most of them have the predictable meaning of ‘small X’. In two instances the suffixed noun has a completely unpredictable, lexicalized meaning, cf. *wódka* ‘vodka’ versus *woda* ‘water’ and *stolek* ‘stool’ versus *stół* ‘table’. In two other cases, the meaning is somewhat lexicalized, cf. *wieczorek* ‘soiree’ versus *wieczór* ‘evening’, and *zbiorek* (e.g. *poezji*) ‘anthology (of e.g. poetry)’ versus *zbiór* ‘collection; harvest’. As to the rest of the listed nouns, they either form diminutives with other suffixes, e.g. *pokoik* ‘room-dim.’, *samochodzik* ‘car-dim.’, or do not have diminutives at all, e.g. *wyбір* ‘choice’, *zachód* ‘West’.

Similar productivity characterizes diminutives derived from nouns of rarer frequencies, cf. *ozdóbka* ‘ornament-dim.’, *potworek* ‘monster-dim.’, whose base nouns rank as 8739-10355 in *Słownik* (1990)), as well as from quite infrequent base nouns, cf. *brzoźka* ‘birch-dim.’, *miodek* (or *miodzik*) ‘honey-dim.’ On the whole, diminutives which

<sup>20</sup> I make my own native speaker’s judgments here, which may be to some extent idiolectal, e.g. I include *narodek* ‘nation-dim.’, but not *prośbka* ‘request-dim.’ Nevertheless, such questionable examples are very few.

are potentially subject to the *o~u* alternation create a fairly sizable class, in which the stem alternation could be maintained, as it is in unsuffixed nouns, by pattern analogy. But it is not, and it seems quite clear that the reason for stem leveling in diminutives relates to the fact that they occur almost only in low frequencies. Hence, there is no salient, memorized “model” which would constitute the base of an alternating pattern and could attract less frequent nouns. Therefore, since pattern analogy may not apply, stem analogy takes place, as is usual in infrequent words.

The approach presented here is consistent with the assumption that there was a stage when the alternation applied within diminutives, as well, and was then eliminated by analogy. Unfortunately, historical evidence concerning the development of diminutives is scarce, especially since the distinction between the *o* vowel and its newly raised variant was not marked in Old Polish orthography. The dictionary of the 16<sup>th</sup> century Polish (*Słownik* 1968-2004) very seldom disambiguates the two vowels in diminutives. But in the case of *ogródek* ‘garden-dim.’, the *o~u* variation is mentioned for the nominative=accusative case, while it is specifically stated that only *u* is found in other cases (i.e. in closed syllables). Similarly, Bańkowski (2000) gives *grodek* (the present form *gródek*) as an earlier diminutive of *gród* ‘fortified town’. (The argument for earlier regular phonology in diminutives is supported indirectly by the data discussed in the following chapter and relating to another alternation which is also leveled in diminutives now and for which there is limited evidence of the earlier alternation.)

Let us now look at the distribution of the diminutives (potentially *o~u* “alternable”) in *Słownik* (1990). Recall from the earlier tables, XI and XIII, that one such noun is found between ranks 1-1002 (*kółko* ‘circle-dim.’) and one between ranks 1003-2009 (*wódka* ‘vodka’). We can conclude that the frequency dictionary confirms the earlier statement about the near non-occurrence of diminutives among high frequency words. Their number stays small even in lower frequencies; among ranks 8739-10355, I have counted only one instance of a true diminutive with the leveled *o/u* (*słówko* ‘word-dim.’), one lexicalized word (*klódka* ‘padlock’, cf. *kloda* ‘log’) and one in which the (*e*)*k* suffix has a nominalizing function (*rozbiórka* ‘demolition’, cf. *rozbiór* ‘partition-nom.’, *rozbioru* ‘partition-gen.’).

The frequency of a diminutive is typically many times lower than the frequency of its base noun. This is illustrated in table XV, which contains the PWN corpus frequencies of the first several nouns of table I (i.e. the highest ranked nouns of *Słownik* 1990) and the frequencies of their corresponding diminutives. Let us also observe that with the exception of the relatively high occurrence of the lexicalized *wódka* ‘vodka’, diminutives have low frequency.<sup>21</sup>

---

<sup>21</sup> Diminutives have slightly higher frequency in particular pragmatic contexts, e.g. while speaking to small children. For some strange sociolinguistic reasons, diminutives are just loved by owners of small grocery stores and boutiques, who sell us *chlebek* ‘bread-dim.’, *bułeczki* ‘rolls-dim.’, *pomidorki* ‘tomatoes-dim.’, *bluzeczki* ‘blouses-dim.’, etc., for which they take *pieniążki* ‘money-dim.’

XV. The PWN corpus frequencies of the *o~u* alternating nouns and their non-alternating diminutives.

base noun (BN) and gloss	BN' occurrences	diminutive	D.'s occurrences
OSOBA 'person'	43007	OSÓBKA	29
WODA 'water'	15519	WÓDKA	1592
SŁOWO 'word'	15419	SŁÓWKO	243
GŁOWA 'head'	14635	GŁÓWKA	401
ROZMOWA 'conversation'	13581	ROZMÓWKA	33
ROLA 'role' <sup>22</sup>	11024	RÓLKA	21
NARÓD 'nation'	5651	NARODEK	1
BÓG 'god'	8917	BOŻEK	101

Low frequency constitutes a solid argument for an analogy-based analysis of diminutives, as does their phonological shape. Let us first concentrate on the feminine and neuter paradigms, in which the direction of analogy is straightforward. Recall from the examples in (15) above that in all declensional cases except for the genitive plural the stem vowel is phonologically predicted to be *u* and this vowel is regularized in diminutives. An analogy-based analysis immediately explains that it should be so, given the extremely low frequency of the genitive plural vis-à-vis joint frequencies of other cases constituting the major *u*-pattern. Even if the genitive is one of the most frequent cases, it is generally much lower in plural, since plurals typically occur less often than singulars. And, naturally, this particular case occurs less often than other frequent cases of the singular, e.g. the nominative and the accusative. Table XVI provides some actual examples of frequencies of WORD and genitive plural forms of several feminine and neuter nouns mentioned earlier. We can see that the percentage of the minor form is really small. It would be impossible to find any lexical exception of a feminine or neuter noun which would generalize the *o* vowel in the diminutive, just as it would be hard to think of any noun in which the use of the genitive plural would be higher than the joint frequencies of all other declensional cases.

Table XVI. The PWN frequencies of WORD and genitive plural word-forms.

WÓDKA 'vodka'	1592: <u>wódek</u> 131 (8,23%)
KÓŁKO 'circle-dim.'	956: <u>kółek</u> 208 (21,76%)
GŁÓWKA 'head-dim.'	401: <u>główek</u> 13 (3,24%)
SŁÓWKO 'word-dim.'	243: <u>słówek</u> 27 (11,11%)
OSÓBKA 'person-dim.'	29: <u>osóbek</u> 0 (0%)
ROZMÓWKA 'conversation-dim.'	33: <u>rozmówek</u> 5 (15,15%)
RÓLKA 'role-dim.'	21: <u>rólek</u> 0 (0%)

In the masculine diminutives, the phonologically predicted *o*-stem is much less frequent than the *u*-variant as far as a number of word-forms in the paradigm is

<sup>22</sup> This word has a homonym meaning 'cultivated land' whose occurrences are included in the count, but the diminutive may be formed only in the meaning of 'role'.

concerned. Recall from (15) that the latter is predicted only in one word-form, which is used as the nominative of animate nouns or the nominative and the accusative of inanimate nouns. However, both these declensional cases are very frequent in Polish. It is not unlikely, especially in the case of inanimate nouns, that the paradigm minor stem pattern can be comparable in size to the paradigm major stem pattern in terms of text occurrence. Whether this happens or not depends largely on the meaning of each individual word, which governs its use in particular semantic and syntactic contexts. For example, the chance of having a frequent *o*-variant increases if a word is used mostly in the singular and in strong syntactic positions, i.e. as a subject or an object (of non-animates). Likewise, the percentage frequency of the *u*-variant increases if a word is a place name and is often used in the locative case, or when it denotes objects often used in the plural, etc. To sum up, the direction of analogy in the case of masculine diminutives is not as unambiguously determined by frequency as in the case of the feminine and neuter genders. This is probably the reason why not all of masculine nouns regularized in the same direction. Most of the masculine diminutives have the *o* vowel, as exemplified in (16a), but there are very few exceptions with the *u* vowel, such as those indicated below in (16b). There is also one case of free variation, given in (16c)<sup>23</sup>, and one case of the lexical split, shown in (16d).

- |       |  |                                     |
|-------|--|-------------------------------------|
| (16a) | bożek ‘god-dim.’   | dołek ‘hole-dim.’                   |
|       | aniołek ‘angel-dim.’   | wzorek ‘pattern-dim.’               |
|       | rowek ‘groove-dim.’  | stworek ‘creature-dim.’             |
|       | rożek ‘horn-dim.’  | utworek ‘work-dim.’                 |
| (16b) | wózek ‘cart-dim.’  | ogródek ‘garden-dim.’               |
|       | kościółek ‘church-dim.’  | gródek ‘fortified town-dim. (obs.)’ |
| (16c) | dziobek~dzióbek ‘beak-dim.’                                    |                                     |
| (16d) | żłobek ‘day care center’ vs żłóbek ‘trough-dim.’ <sup>24</sup> | (cf. żłób ‘trough, manger’)         |

The left-hand column of Table XVII provides the frequencies of WORDs and nominative (=accusative) word-forms of sample diminutives, including those which leveled the *u* vowel. If the leveling is triggered by frequency criteria within the paradigm, we expect the nom. (=acc.) minor form to reach more than 50% in the XVIIa examples and less than 50% in the XVIIb examples. But the data do not show such a correlation or any other coherent pattern which would account for a difference between the two classes. At the same time the data support the previous statement that the percentage use of the nominative (=acc.) varies greatly depending on the given lexical item, which makes a generalization impossible. Let us also observe that many diminutives have very low frequency and as such, must be retrievable by pattern-analogy and not by stem-analogy within their own paradigm. In other words, none of the word-forms of very rare frequency diminutives would be salient enough to create a Base for analogy. But this creates a vicious circle: none of the stem variants appears as a salient Base for analogy within a paradigm, and pattern analogy is impossible because there is too much variation within the category. The situation in the masculine gender diminutives is thus quite

<sup>23</sup> Another example is *stópka~stopka* ‘foot-dim.’, but this is an exceptional case of the alternation before a phonemic voiceless consonant.

<sup>24</sup> The word is used in the religious context as the “cradle” of little Jesus.

different from that of feminine and neuter ones, in which the same *u*-stem variant was consistently more frequent and consequently – salient, in spite of infrequent occurrences of individual word-forms.

It might be suggested that the analogical unmarked *o*-pattern for masculine diminutives is triggered by the higher frequency of this form in non-diminutive nouns. However, the relevant figures, included in the right hand column of Table XVII, show that this variant is generally only slightly more frequent than the nominative *u*-variant. And the exceptional data of XVIIb remain equally unexplained, since the *u*-variant is by no means more frequent.

Table XVII. The PWN frequencies of WORD and the nom. (=acc.) word-form.

a/ *o*-leveled

DOŁEK 306: <u>dołek</u> 88 (28,76%)	DÓŁ 3882: <u>dół</u> 2292 (59,04%) ‘hole’
BOŻEK 101: <u>bożek</u> 49 (48,51%)	BÓG 8917: <u>bóg</u> 2449 (27,46%) ‘god’
ROŻEK 78: <u>rożek</u> 39 (50%)	RÓG 1250: <u>róg</u> 267 (21,36%) ‘horn’
ROWEK 49: <u>rowek</u> 6 (12,24%)	RÓW 421: <u>rów</u> 70 (16,63%) ‘groove’
WZOREK 39: <u>wzorek</u> 9 (23,08%)	WZÓR 3660: <u>wzór</u> 1077 (29,43%) ‘pattern’
STWOREK 13: <u>stworek</u> 0 (0%)	STWÓR 164: <u>stwór</u> 55 (33,54%) ‘creature’
UTWOREK 0: <u>utworek</u> 0	UTWÓR 3322: <u>utwór</u> 588 (17,70%) ‘work’

b/ *u*-leveled

OGRÓDEK 755: <u>ogródek</u> 224 (29,69%)	OGRÓD 2212: <u>ogród</u> 505 (22,83%) ‘garden’
WÓZEK 731: <u>wózek</u> 223 (30,51%)	WÓZ 1585: <u>wóz</u> 426 (19,26%) ‘cart’
KOŚCIÓŁEK 157: <u>kościółek</u> 60 (38,22%)	KOŚCIÓŁ 9656: <u>kościół</u> 3053 (31,62%) ‘church’
GRÓDEK 28: <u>gródek</u> 10 (35,71%)	GRÓD 355: <u>gród</u> 119 (33,52%) ‘fortified town (obs.)’

An alternative analysis may assume that the *o*-diminutives are leveled towards the “unmarked” nominative singular (cf. Kraska-Szlenk 1995/2003). In the psychological sense, this case certainly has a privileged status, since it occurs in a syntactically strong position of the subject, often sentence initially, appears as a dictionary form, etc. However, there does not seem to be conclusive empirical evidence that a “psychologically prominent” rather than the most frequent form may constitute a Base for analogy. It is likely though, that such “psychological prominence” comes into play when frequency criteria are not decisive, just as in the case described here. But how can be the exceptions in XVIIb accounted for? Possibly, it is not coincidental that all of these nouns denote locations and often occur with the prepositions *w* ‘in’ (used with the locative), *do* ‘to, into’ or *z* ‘from’ (both used with the genitive), which highly increases the use of the closed-syllable stem variant, phonologically predicted to contain the *u* vowel, cf. (*rzeczy*) *w wózku* ‘(things) in the cart’, (*włożyć*) *do wózka* ‘(put) into the cart’, (*wyjąć*) *z wózka* ‘(take out) from the cart’, (*kwiatki*) *w ogródku* ‘(flowers) in the garden-dim.’, (*iść*) *do ogródka* ‘(go) to the garden-dim.’, (*kwiatki*) *z ogródka* ‘(flowers) from the garden’.

Perhaps for nouns like these, psychologically “prominent” forms are those expressing locations.

To conclude, even though the mechanism of analogy is not entirely clear in the case of masculine diminutives, it seems that the lack of a uniform pattern is due to the fact that frequency criteria are unable to unambiguously determine the Base for analogy, which leaves room for variation.

### 3.4. Double diminutives

The Polish *ek* suffix may occur in a reduplicated form, as it often happens in diminutives cross-linguistically (cf. Spanish *gato* ‘cat’, *gatito* ‘cat-dim.’, *gatitito* ‘cat-dim.-dim.’). However, the class of double diminutives is pragmatically restricted, so that not every noun that may occur with the single *ek* suffix sounds acceptable with the reduplicated suffix, cf. *aniołek* ‘angel-dim.’ and *anioleczek* ‘angel-dim.-dim.’,<sup>25</sup> but *bożek* ‘god-dim., idol’, \**bożeczek* ‘god-dim.-dim.’ Out of 25 sample nouns mentioned earlier, which occur in the diminutive form, only ten make sensible double diminutives, including two which are lexicalized in their “first” diminutive form, cf. (with the PWN corpus WORD frequencies indicated in brackets): *wódeczka* (41) ‘vodka(lexicalized)-dim.’, *głoweczka* (0) ‘head-dim.-dim.’, *słoweczko* (7) ‘word-dim.-dim.’, *nóżeczka* (0) ‘leg-dim.-dim.’, *kółeczko* (27) ‘circle-dim.-dim.’, *wózeczek* (12) ‘cart-dim.-dim.’, *stołeczek* (41) ‘stool (lexicalized)-dim.’, *doleczek* (23) ‘hole-dim.-dim.; dimple’, *króweczka* (0) ‘cow-dim.-dim.’, *ogródeczek* (0) ‘garden-dim.-dim.’.

While single-suffixed diminutives have low text occurrence, double diminutives are extremely rare, which is already seen from the PWN corpus figures shown above. In *Słownik* (1990), none of them appeared among the most frequent words (up to rank 2009) and there were only two double diminutives among lower frequency words (ranks 8739-10355), but none of them with the *o~u* “alternable” stem. Even in texts in which diminutives occur more often than usually, e.g. in children’s stories, double diminutives are hard to find, cf. table XVIII (after Kraska-Szlenk 1999a).

Table XVIII. Ratios of Polish nouns in a sample text (children’s stories).

noun type	number of occurrences	percentage of occurrence
non-diminutive	1269	85.5% (of all nouns)
<i>ek</i> -suffixed diminutive	173	11.7%
double diminutive	0	0%

(NB: The percentages do not add up to 100, because of the occurrence of nouns with diminutive suffixes other than *ek*.)

The extremely low frequency of double diminutives explains why the stem leveling is oriented toward a Base from outside the paradigm and copies the stem vowel of the Base diminutive, as illustrated in (17) below.

<sup>25</sup> The *ek-ek* suffix is subject to palatalization, giving the surface [eʃek]. I leave out this issue here.

(17a) The singular paradigm of the double diminutives of *o~u* nouns

case	masculine	feminine	neuter
nom sg	wózeczek, dołeczek	głoweczka	słoweczko
gen sg	wózeczka, dołeczka	głoweczki	słoweczka
dat sg	wózeczкови, dołeczкови	głoweczce	słoweczku
acc sg	wózeczek, dołeczek	głoweczkę	słoweczko
instr sg	wózeczkiem, dołeczkiem	głoweczką	słoweczkiem
loc sg	wózeczku, dołeczku	głoweczką	słoweczku
voc sg	wózeczku, dołeczku 'cart', 'hole, dimple'	głoweczko 'head'	słoweczko 'word'

(17b) The plural paradigm of the double diminutives of *o~u* nouns

case	masculine	feminine	neuter
nom pl	wózeczki, dołeczki	głoweczki	słoweczka
gen pl	wózeczków, dołeczków	głoweczek	słoweczek
dat pl	wózeczkom, dołeczkom	głoweczkom	słoweczkom
acc pl	wózeczki, dołeczki	głoweczki	słoweczka
instr pl	wózeczkami, dołeczkami	głoweczkami	słoweczkami
loc pl	wózeczkach, dołeczkach	głoweczkach	słoweczkach
voc pl	wózeczki, dołeczki 'carts', 'holes, dimples'	głoweczki 'heads'	słoweczka 'words'

Let us observe that in double diminutives, the stem vowel always occurs in an open syllable, but it happens to be *u* in all feminine and neuter, as well as in a few exceptional masculine nouns. Consequently, a purely phonological analysis of the data is untenable. It is hard to sustain any cyclic analysis, either (as e.g. Laskowski 1975, Szypra 1992), because in the inner cycle the nouns of all three genders have the same syllable structure (open or close, depending on analysis, cf. (18) and (19) below), hence they should ultimately have the same vowel on the surface. (An additional problem would be the existence of exceptions with the *u* stem vowel among the masculine nouns.) I illustrate the problem in sketchy cyclic analyses below, which attempt to be maximally neutral with respect to theoretical assumptions. I use the capital *O* for the underlying "alternable" stem vowel and I use (*e*) for the *e~∅* alternating suffixal vowel. The crucial difference between the two analyses is ordering of the (*e*)-vocalization as *e* and the realization of *O* as *o* or *u* depending on syllable position. But under each analysis only one of the paradigms is derived correctly and one is always flawed. That is because there are no structural differences between the word-forms of the masculine paradigm and the feminine/neuter paradigm (e.g. the masculine nom. sg. has an identical structure as the feminine gen. pl. and the masculine gen. sg. has an identical structure as the feminine nom. sg.), but their surface stem vowels are different.

## (18) Hypothetical type I cyclic analysis of double diminutives

	masc. nom. sg.	fem. gen. pl.	derivational stage
a/	[[dOw-(e)k]-(e)k] [[dOwek]-(e)k] [[dɔwek]-(e)k] [dɔwetʃek] 'dimple'	[[gwOv-(e)k]-(e)k] [[gwOvek]-(e)k] [[gwɔvek]-(e)k] *[gwɔvetʃek] 'head-dim.-dim.' (correct: [gwuvetʃek])	underlying <i>e</i> -Vocalization <i>o</i> in open syllable other rules gloss
b/	[[dOw-(e)k]-(e)ka] [[dOwek]-(e)ka] [[dɔwek]-(e)ka] [dɔwetʃka] 'dimple'	[[gwOv-(e)k]-(e)ka] [[gwOvek]-(e)ka] [[gwɔvek]-(e)ka] *[gwɔvetʃka] 'head-dim.-dim.' (correct: [gwuvetʃka])	underlying <i>e</i> -Vocalization <i>o</i> in open syllable other rules gloss

## (19) Hypothetical type II cyclic analysis of double diminutives

	masc. nom. sg.	fem. gen. pl.	derivational stage
a/	[[dOw-(e)k]-(e)k] [[duw(e)k]-(e)k] [[duwek]-(e)k] *[duwetʃek] 'dimple' (correct: [dɔwetʃek])	[[gwOv-(e)k]-(e)k] [[gwuv(e)k]-(e)k] [[gwuvek]-(e)k] [gwuvetʃek] 'head-dim.-dim.'	underlying <i>u</i> in closed syllable <i>e</i> -Vocalization other rules gloss
b/	[[dOw-(e)k]-(e)ka] [[duw(e)k]-(e)ka] [[duwek]-(e)ka] *[duwetʃka] 'dimple' (correct: [dɔwetʃka])	[[gwOv-(e)k]-(e)ka] [[gwuv(e)k]-(e)ka] [[gwovek]-(e)ka] [gwuvetʃka] 'head-dim.-dim.'	underlying <i>u</i> in closed syllable <i>e</i> -Vocalization other rules gloss

Another theoretically possible analysis may assume that double diminutives are derived from surface “first” diminutive forms and that is why they share with them the same stem vowel. Such an analysis is in essence very similar to analogy-based explanation, except that it involves derivational steps. The simplest and most straightforward way of deriving double diminutives is to assume that the second instance of the *(e)k* (in its phonologized form, or *eʃ* in its surface form) suffix is infixes directly after the root and before the first instance of the diminutive suffix, as illustrated in (20a).

If the second suffix were to appear after the first one, we would face a problem with choosing the adequate base for suffixation, because in the “first” diminutives the stem occurs in two variants, cf. masculine nouns: the nominative singular {dɔwek} and the genitive singular {dɔwk}a, and feminine nouns: the nominative singular {gwuvk}a and the genitive plural {gwuvek}. It would have to be somehow stipulated that the longer stem variant, {dɔwek} and {gwuvek}, would have to be chosen, but that would be very odd in the case of the feminine/neuter declension, since this variant occurs only in the genitive plural. The direct infixation approach also accounts for the possibility of using the *eff* suffix in a recursive manner, e.g. *doleczeczek*, *doleczeczeczek*, etc., which is far from the formal language use, but freely happens in all kinds of jocular situations, especially with children. An OT equivalent of the *eff*-infixation analysis involves clusters of correspondence constraints, as sketched in (20b). Cluster A includes all intraparadigmatic correspondence constraints, which have a high ranking, motivated by rare occurrence of double diminutives. The minimal violation of stem correspondence, viz. *doleczek-*, *doleczk-*, is enforced by the alternation of the *ek* suffix, discussed in the following chapter 4. Cor-B includes correspondence relations between word-forms of double diminutives and their Bases i.e. the relevant forms with a single suffix, which reflects the intuition of double diminutives as being derived from them. The correspondence here is partially violated (with respect to contiguity) due to the infixation of *eff*, enforced by a highly ranked morphological constraint (M-DbIDim). The ranking of this constraint below Cor-A, enforces the violation of its “suffixness” and moves it into an infix position. I assume that this is a new synchronic reinterpretation of a historical suffix, which originally followed the first instance of a diminutive suffix. All three constraints mentioned here dominate the *o~u* “alternation” constraint discussed in the next section, as well as all other correspondence constraints between stem-sharing forms.<sup>26</sup>

(20a) A derivational analysis with *eff*-infixation

m.nom.sg.	m.gen.sg.	f.nom.sg.	f.gen.pl.	derivational step
dɔwek	dɔwka	gwuvka	gwuvek	first diminutive
dɔw{eff}ek	dɔw{eff}ka	gwuv{eff}ka	gwuv{eff}ek	<i>eff</i> -infixation
dɔweʃfɛk	dɔweʃfka	gwuveʃfka	gwuveʃfɛk	surface outputs

---

<sup>26</sup> The analysis presented here is not the only one possible. For example, a more “standard” analysis based on recursive suffixation may appeal to syllable structure conditions in choosing a base in double diminutives, as it will be proposed for another class of diminutives in chapter 4.

(20b) A correspondence-based analysis with *eff*-infixation

Ranking: Cor-A >> M-Dbldim >> Cor-B

Cor-A {*doleczek: dołeczka, dołeczka: dołeczek* etc.}

Cor-B {*doleczek: dołek, dołeczka: dołka*, etc.}

M-Dbldim: must contain suffix *eff*

To conclude, the data presented in this section have illustrated several important aspects of stem and pattern analogy in relation to frequency in language use. The fact of the stable character of the *o~u* alternation in base nouns is supported by a large size of the class, as well as the existence of a significant group of lexical items of high frequency within that class. The elimination of the alternation within diminutives by stem-analogy within their paradigms directly relates to their rare occurrence. Finally, the extremely infrequent double diminutives are so hard to retrieve that a salient Base for their leveling must come from outside their paradigm.

### 3.5. An Optimality Theoretic analysis of the *o~u* alternation

An OT analysis of the data included in this chapter involves correspondence constraints correctly ranked with respect to a relatively general constraint against the vowel *o* before a [voiced, -nasal] coda (abbreviated as \*oD(C)]<sub>σ</sub>). A large class of nouns obey this constraint and simultaneously bring about a stem final *o~u* paradigmatic alternation which creates a strong recognizable pattern. This templatic face of the originally phonological constraint allows for words of rarer occurrence to follow more salient ones in forming a unified class, provided they share a similar morphophonological make-up. The latter includes appropriate phonological features, but also a stem final position, because this is the morphological environment of the alternation present in many salient, frequent nouns. Therefore, low frequency words as e.g. the above-mentioned *glóg* ‘hawthorn’, *znój* ‘toil’ or *płozą* ‘runner (of sleigh)’ can have their correspondence constraints placed together with those of “core” frequent words (e.g. *wóz* ‘car’, *głowa* ‘head’), as exemplified in cluster *A* in (21) below. This strategy is not available to diminutives, which are also infrequent, because class *A* is not accessible to them for structural reasons. Since they do not have salient representatives of their own able to create an alternating pattern in the position away from the end of the stem, they have to disobey the phonological constraint and aim for stem correspondence, to which I will come momentarily. In the meantime, let us acknowledge a smaller number of non-alternating nouns with the fixed vowel *o*, e.g. *kod* ‘code’, *metoda* ‘method’, which form an “exceptional”<sup>27</sup> class *B*.

---

<sup>27</sup> The notion „exceptional” does not have a formal status and directly relates only to the category size (cf. Kraska-Szlenk 1999b).

(21) An OT analysis of different lexical classes

Ranking: Cor-B >> \*oD(C)]<sub>σ</sub> >> Cor-A

Cor-A {wóz: wozu, głowa: głów, głóg: głogu, płoza: plóz etc.}

Cor-B {kod: kodu, metoda: metod etc.}

Returning to diminutives, I will first concentrate on a more evident case of feminine and neuter genders. Recall that in these diminutives, the leveled stem form is predicted by simple majority criteria, which in an OT analysis follows directly from usage evaluation *via* extended correspondence constraints evaluated pair-wise, as in McCarthy's (2005) approach. Intraparadigmatic diminutive constraints (cluster *Dim* below) must dominate correspondence constraints between diminutives and their base nouns (cluster *Dim/A*) to guarantee that the Base of the diminutive will not be triggered by the majority constraints of the latter. This contrasts with the behavior of double diminutives, too rare to have a Base of their own and dependent through correspondence (Cor-DblDim/Dim) on their next of kin, as illustrated in (22) below. I assume that intraparadigmatic double diminutive constraints (Cor-DblDim) are the highest in the ranking and trigger Cor-DblDim/Dim.

(22) An OT analysis of diminutives

Ranking: Cor-DblDim >> Cor-DblDim/Dim >> Cor-Dim >> Cor-Dim/A >> \*oD(C)]<sub>σ</sub>

Cor-Dim {główka: główki, główka: główek, główek: główka etc.}

Cor-Dim/A {główka: głowa, główka: głowy, główka: głów etc.}

Cor-DblDim/Dim {głoweczka: główka, główeczek: główek etc.}

Cor-DblDim {głoweczka: główeczki, główeczek: główeczka etc.}

The lexical split between masculine diminutives can be attributed to the split within their correspondence constraints, reflecting the suggested earlier “psychological prominence” of the nominative case for most nouns and of the locative/genitive (i.e. cases expressing locations) for some nouns denoting places. (Recall that this “extraordinary” strategy comes into play only because frequency criteria are not decisive in the case of masculine nouns.) Prominent cases are consequently promoted as Bases over the rest of intraparadigmatic correspondence constraints of diminutives, as illustrated in (23). (The rest of the analysis, as similar to that for feminine and neuter nouns, is not included.)

(23) Ranking: Cor-X-Dim/Dim-nom, Cor-Y-Dim/Dim-loc(gen) >> Cor-Dim

Cor-X-Dim/Dim-nom {dolka: dołek, dolki: dołek, bożka: bożek etc.}

Cor-Y-Dim/Dim-loc(gen) {wózek: wózka, wózek: wózku, ogródek: ogródka etc.}

## CHAPTER 4

### The *e~∅* alternation in Polish nouns

#### 4.1. The historical background and the current scope of the alternation

The *e~∅* alternation, which is the subject of this chapter, differs in character from both alternations previously discussed. It does not show the steadiness of the *o~u* alternation and, unlike the *o~e* alternation, it is not much susceptible to stem analogy. On the contrary, it creates a strong morphophonemic pattern and may exert pattern analogy on new lexical items.

The *e~∅* alternation poses a long-standing problem in the grammar of Polish. No other alternation has received so much attention from phonologists and sprouted so many different treatments (cf. e.g. Rowicka 1999:170-179 for a concise review). Underlying phonological representations of the “ghost” vowels include, among others, abstract vowels (e.g. Gussmann 1980, Jarosz 2005, Rubach 1984, 1986), floating features (e.g. Rubach and Booij 1990), empty nuclei (e.g. Rowicka 1999, Spencer 1986) or surface reflexes of underlying syllabic consonants (Piotrowski *et al.* 1992). Other proposals assume that “ghost” vowels have epenthetic nature and emerge on the surface as a result of syllabification in particular morphophonological environments (e.g. Laskowski 1975, Piotrowski 1992).

The present *e~∅* alternation is rooted in the historical process which eliminated extra short high vowels, the so-called “yers”, from the phonological system of Polish. Sporadic yer deletion occurred at various times, including even the proto-Slavic period (cf. *\*kьto > kto* ‘who’), but as a regular process it took place in Polish around the 11<sup>th</sup> century. Original yers, which occurred in two phonemic variants, the back “hard” *ь* and the front “soft” *ѣ*, atrophied in prosodically weak positions and surfaced as the vowel *e* in prosodically strong positions. In addition, the soft yer caused palatalization of the preceding consonant and this feature remained in the case of the yer’s deletion.

The yer deletion/vocalization process had a rhythmic character. A yer was realized when another yer followed in the next syllable, which was in turn deleted. Deletion of yers also took place word-finally and before a full vowel, as the examples in (24) illustrate. For practical reasons, it is convenient to apply all the rules iteratively from the end of the word, or, in an OT non-derivational style, to subordinate the rhythmic constrain to the higher constraints of word-final and before-the-vowel deletion.

#### (24) The regular development of yers

Early Polish		Old Polish	contemporary form+gloss
ѣВѢТСѢ	>	ʃv <sup>i</sup> ets	ʃefts (ort. szewc) ‘shoe-maker-nom.’
ѣВѢТСА	>	ʃeftsa	ʃeftsa (ort. szewca) ‘shoe-maker-gen.’
рѢСѢКѢ	>	psek	p <sup>i</sup> esek (ort. piesek) ‘dog-dim.-nom.’
рѢСѢКА	>	p <sup>i</sup> eska	p <sup>i</sup> eska (ort. pieska) ‘dog-dim.-gen.’

The historical principles of yers' vocalization/deletion are of little use for a synchronic description of the "ghost" vowels. The rhythmic rule is hardly detected, because there are no sufficiently long sequences of yers, unless the word-final yer is posited for the zero inflectional ending. But this hypothetical yer (always realized as phonological zero) would not do any other job but make the rhythmic rule operate. In other cases of long sequences of earlier yers, original forms were often leveled by stem or pattern-analogy (cf. the earlier examples in (24)), which makes the rhythmic rule indecipherable. Therefore, from the synchronic perspective, the conditions of the  $e\sim\emptyset$  alternation are much simpler and can be stated as: a "ghost" vowel deletes before a vowel in the following syllable; otherwise it is realized as  $e$ . The historical distinction between the soft and hard yers is neutralized, if a palatal feature is specified on the consonant, which is typically assumed in most analyses of the issue, especially since palatalized consonants have a phonemic status in Polish (cf. chapter 2). Thus for example, an earlier word as  $*vbsb$  'village' will have a modern phonological interpretation as  $v^j(e)\epsilon$ , where  $(e)$  denotes an  $e\sim\emptyset$  alternating vowel (but it is up to a phonologist how  $(e)$  is actually understood).

Before the disappearance of yers, Polish was predominantly an open-syllable language. After the deletion, the situation changed drastically, because the process was completely irrespective of any syllable structure constraints. As a result, the language acquired clusters which heavily violate sonority hierarchy and which often look as multiple well-formed onsets.<sup>28</sup> Word-initially, there are many such "bad" consonant sequences in Polish, which happen to occur in one or two lexical items, while a similar or even "better" sequence is not found. Similarly, haphazardous consonantal sequences are found word-medially and word-finally, which is mostly due to the idiosyncratic yers deletion in suffixes, but also to analogical stem-leveling (as in *szewc*, shown earlier in (24)). Examples in (25-27) below provide a brief illustration of remarkable clusters found in initial, medial and final positions, respectively.

(25) Word-initial clusters:

[tskl] ckliwy 'sentimental'	[pxw] pchła 'louse'
[tʃtʃ] czczy 'meaningless'	[krt] krtań 'larynx'
[mgw] mgła 'fog'	[brn] brnąć 'to wade'
[klɲ] klnie '(he) swears'	[rʒn] rżnąć 'to saw'
[lɕn] lśnić 'to shine'	[drgn] drgnąć 'to shudder'
[zɖbw] źdźbło 'stalk (of grass)'	[pstr] pstry 'gaudy'

<sup>28</sup> The phonological treatment of such sequences and the issue of the Polish syllable are far too complex to be treated here, cf. e.g. Kraska-Szlenk 2003:7-10, Rowicka 1999:179-187 for discussions of various proposals and the references. In the English language literature, an extensive list of Polish clusters is given in Rowicka (1999:309-344).

(26) Word-medial clusters:

[mbrn] krnąbrny ‘unruly’	[snk] piosnka ‘song’
[mstf] kłamstwo ‘lie’	[ndrk] mędrkować ‘to play the wise guy’
[ndrn] jędrny ‘firm’	[ntʃk] garneczka ‘pot-dim.-gen.’
[nktsj] funkcja ‘function’	[ntpl] wątpliwy ‘questionable’
[rskn] parsknąć ‘to snort’	[rʃtʃk] zmarszczka ‘winkle’

(27) Word-final clusters:

[sf] nazw ‘names-gen.’	[ʃx] zmierzch ‘twilight’
[dl~tl] módl (się) ‘pray-imp.’	[ɕm] taśm ‘tapes-gen.’
[ɕp] prósb ‘requests-gen.’	[wm] hełm ‘helmet’ę
[tʃp] liczb ‘numbers-gen.’	[wtɕ] żółć ‘bile’
[xtr] blichtr ‘glare’	[jstf] podobieństw ‘similarities-gen.’
[jstf] zabójstw ‘murders-gen.’	[mpstf] przestępstw ‘crimes-gen.’

The existence of clusters as above (and many others equally remarkable) clearly points out that phonotactic constraints are not highly ranked in Polish. However, it is worthwhile to observe that, even though there are many types of such “strange” consonantal sequences, they occur in few lexical items, and, consequently, are of no value as possible extending patterns. For example, all clusters shown in (25-27) are found uniquely in the mentioned lexemes (stems) and in no others. Similarly, certain sequences occur only in a particular morphophonemic context, as e.g. in the genitive plural of nouns suffixed with *-stwo*. Therefore, in spite of the apparent enormous tolerance of Polish for all kinds of onsets and codas, there are areas in the language where syllable structure constraints do play a role, which I will come to later in this chapter.

In addition to the remnants of the earlier yers and present “ghost” vowels, Polish has a non-alternating *e* vowel, as well as phonological zero, which occur in identical environments. Actually, the data of masculine nouns such as those in (28), including the same *t(e)r* sequence, could give a phonologist a real headache, since in each of the four examples the sequence has different realizations in the same environment. The *tr* cluster is never broken in the base noun and its diminutive in (28a); it appears as *tr* in the base noun and as *ter* in the diminutive in (28b); *tr* occurs only in the genitive of the base noun in (28c), otherwise it is *ter*; and *ter* is found in all four forms of (28d). Luckily, the alternation in (28b) is rather exceptional, and the remaining examples illustrate typical cases: a *CC* non-alternating cluster in (28a), a cluster with a “ghost” vowel in (28c), and a *CeC* non-alternating sequence in (28d). (NB all three words are loanwords, but this fact is quite irrelevant.)

## (28) Alternating and non-alternating paradigms

	nominative	genitive	diminutive-nom.	diminutive-gen.	gloss
a/	Pi <u>otr</u>	Pi <u>otra</u>	Pi <u>otrek</u>	Pi <u>otrka</u>	‘Peter’
b/	wi <u>atr</u>	wi <u>atru</u>	wi <u>aterek</u>	wi <u>aterku</u>	‘wind’
c/	sw <u>eter</u>	sw <u>etra</u>	sw <u>eterek</u>	sw <u>eterka</u>	‘sweater’
d/	boh <u>ater</u>	boh <u>atera</u>	boh <u>aterek</u>	boh <u>aterka</u>	‘hero’

I will first concentrate on the base nouns and postpone the analysis of diminutives until sections 4.2.3 and 4.3.3. The non-alternating nouns of the types illustrated in (28a) and (28d) preserve the respective *CC* or *CeC* stem sequence throughout the declensional paradigm in the singular and plural. The paradigms of alternating nouns of all three genders are shown in (29) below. We can see the generalization familiar from chapter 3: the *CeC* variant (underlined) is found only before the zero case endings, i.e. in the nominative (equal to the accusative in inanimates) of masculine nouns and in the genitive plural of feminine and neuter nouns. As already argued in chapter 3, the latter case appears as particularly infrequent in terms of text occurrence.

(29a) The singular paradigm of the *e~∅* alternating nouns

case	masculine	feminine	neuter
nom. sg.	<u>sweter</u> , pies	matka	okno
gen. sg.	swetra, psa	matki	okna
dat. sg.	swetrowi, psu	matce	oknu
acc. sg.	<u>sweter</u> , psa	matkę	okno
instr. sg.	swetrem, psem	matką	oknem
loc. sg.	swetrze, psie	matce	oknie
voc. sg.	swetrze, psie ‘sweater’, ‘dog’	matko ‘mother’	okno ‘window’

(29b) The plural paradigm of the *e~∅* alternating nouns

case	masculine	feminine	neuter
nom. pl.	swetry, psy	matki	okna
gen. pl.	swetrów, psów	<u>matek</u>	<u>okien</u>
dat. pl.	swetrom, psom	matkom	oknom
acc. pl.	swetry, psy	matki	okna
instr. pl.	swetrami, psami	matkami	oknami
loc. pl.	swetrach, psach	matkach	oknach
voc. pl.	swetry, psy ‘sweaters’, ‘dogs’	matki ‘mothers’	okna ‘windows’

## 4.2. The $e\sim\emptyset$ alternation in masculine nouns

### 4.2.1. The distribution of alternating masculine nouns within the lexicon

Using *Słownik* (1990) as the source, I will look at the distribution of the  $e\sim\emptyset$  alternating nouns vis-à-vis the non-alternating ones with the stem-final fixed  $CC$  or  $CeC$  sequence. In this section, I will consider masculine nouns, postponing the discussion of feminines and neuters until section 4.3. It should be noted though that a small number of feminine nouns ending in a consonant in the nominative singular are grouped together with masculine nouns, because they pattern with them with respect to the  $e\sim\emptyset$  alternation, cf. nom. *krew*, gen. *krwi* etc. For the same reason, the few masculine nouns ending in  $-a$  in nom. sg., as e.g. *mężczyzna* ‘man’, will be grouped with feminine nouns.

Table XIX contains all the  $e\sim\emptyset$  alternating nouns extracted from first thousand words of the dictionary. In most cases the cluster in the  $\emptyset$ -stem variant consists of only two consonants, but occasionally a sequence of three consonants (underlined> appears on the surface in the non-nominative forms. Several loanwords (marked as “loan”) appear in this group. In some of them, the  $e\sim\emptyset$  pattern analogy affected the words which originally ended in a  $-CC$  cluster, breaking the sequence in the nominative singular by the epenthetic  $e$ , as in all German borrowings with the final  $*-ng$ , cf. *warunek* ‘condition’, *gatunek* ‘kind’, *rynek* ‘market’. The pronunciation with the final  $-nk$  is attested in earlier texts, cf. the contemporary *wizerunek* ‘image’, but *wizerunk* in the 16<sup>th</sup> century (Długosz-Kurczabowa and Dubisz 1998:99). The original German suffix appears also with native stems as a borrowed suffix, e.g. *budynek* ‘building’ in Table XIX, or *kierunek* ‘direction’. It should be noted that the vowel  $e$  is not the phonologically unmarked epenthetic vowel in Polish (it is  $i/i$  that functions in this way), which clearly shows that the vowel insertion here is triggered by the adjustment to the morphophonemic pattern. The same epenthetic  $e$  presumably emerged in the native noun *ogień* ‘fire’ usually reconstructed without the internal yer, which should have become in Polish *ogń* (but see Bańkowski 2000 v.2:394 for a different opinion). In some other loanwords, the original  $CeC$  sequence was reduced to  $CC$  before a declensional ending, cf. *minister* (nom.), *ministra* (gen.) ‘minister’ and *handel* (nom.), *handlu* (gen.) ‘trade’. Although both these strategies are productive, the latter is much more common, which can be probably attributed to the fact that the languages Polish borrows from usually have more restricted syllable structure, so if they tolerate a particular word-final cluster, Polish usually does too; hence there is no need for epenthesis.

Table XIX. The  $e\sim\emptyset$  alternating nouns of masculine gender (unless marked otherwise), ordered according to their rank, as they appear in the first 1000 word-list.

word (nom.) + gloss	rank	frequency
dzień ‘day’	63	718
związek ‘union’	141	356
warunek ‘condition’ (loan)	148	340
minister ‘minister’ (loan)	201	260
koniec ‘end’	208-214	249
środek ‘center’	215	248

stopień ‘degree’	220-223	242
ojciec ‘father’ <sup>29</sup>	229	238
stosunek ‘relation’	251-252	220
członek ‘member’	254-257	218
wypadek ‘accident’	258-260	217
wieś ‘village’	264	212
kierunek ‘direction’	285-287	194
przypadek ‘incident’	379-381	151
wniosek ‘conclusion’	387-391	148
mieszkaniec ‘inhabitant’	428-431	139
statek ‘ship’	450-451	134
początek ‘beginning’	452-454	133
ośrodek ‘center’	522-527	116
porządek ‘order’	652-660	96
skutek ‘result’	670-675	94
<u>handel</u> ‘trade’ (loan)	676-684	93
chłopiec ‘boy’	709-717	89
Niemiec ‘German’	709-717	89
wysiłek ‘effort’	718-729	88
<u>krw</u> (f.) ‘blood’	730-735	87
obowiązek ‘obligation’	736-749	86
budynek ‘building’	765-778	83
piątek ‘Friday’	779-787	82
czwartek ‘Thursday’	811-819	79
ogień ‘fire’	904-913	71
gatunek ‘kind’ (loan)	982-1002	66
odcinek ‘segment’	982-1002	66
palec ‘finger’	982-1002	66
rynek ‘market’ (loan)	982-1002	66

Let us briefly summarize the data of Table XIX with respect to the type of the alternating *CeC~CCV* final sequence (I will use the symbol # for the word-initial position). Out of 35 words undergoing the alternation, 22 have [k] as the second consonant; specifically, we find [tek~tkV] (9 cases), [nek~nkV] (8), [sek~skV] (3), [wek~wkV] (1) and [rték~rtkV] (1). In six words, the second *C* is the affricate [ts]; we find: [nets~ntsV] (2), [jets~jtsV] (1), [lets~ltsV] (1), [m<sup>l</sup>ets~mtsV] (1) and [p<sup>l</sup>ets~ptsV] (1). In three cases the second *C* is [ɲ], namely: [#d<sup>ɛ</sup>ɲ~#dɲV] (1), [g<sup>l</sup>ɛɲ~gɲV] (1) and [p<sup>l</sup>ɛɲ~pɲV] (1). In each of the remaining four words, a different cluster appears, such as: [#v<sup>l</sup>ɛç~#fçV], [ster~strV], [ndel~ndlV] and [#kref~#krf<sup>o</sup>V].

Table XX contains 29 nouns with the non-alternating stem-final *CeC* sequence found in the same range of the first thousand. Sequences that would create three-consonantal clusters under the *e*-deletion in the pre-vocalic environment are underlined.

<sup>29</sup> The cluster is idiosyncratically simplified in other cases, cf. *ojca* (gen.), *ojcu* (dat.) etc.

In addition, a single asterisk indicates that the potential *e*-deletion would create a cluster (of two or more consonants) that is not found in that particular morphological environment, and two asterisks indicate that such a hypothetical medial sequence does not occur in Polish at all. It can be seen that all types of *CeC* sequences found in this group are complementary with respect to the *CeC* sequences of the alternating nouns from Table XIX. Moreover, with the exception of only one borrowing discussed below, all words of this category do not permit *e*-deletion either by purely phonotactic or morphophonological criteria. We can assume that the former rule out medial sequences such as *krs*, *blm*, *cs* etc. in general, making *e*-deletion impossible for all double-starred words. In the case of single-starred words, sequences which would arise under the hypothetical *e*-deletion are found in words such as e.g. *cle* ‘customs duty-loc.’, *getto* ‘ghetto’, *sarna* ‘deer (doe)’, *warga* ‘lip’ etc., but are not found among the *e*~∅ alternating masculine nouns. Only the loanword *szef* ‘boss’ has an alternating homonym, cf. *szew* [ʃɛf] ‘seam-nom.’, *szwu* [ʃfu] ‘seam-gen.’ versus *szefa* [ʃɛfa] ‘boss-gen.’

Table XX. Masculine nouns (unless marked otherwise) with non-alternating *CeC*, ordered according to their rank, as they appear in the first 1000 word-list.

word (nom.) + gloss	rank	frequency
** <u>okres</u> ‘period’	145-146	346
** <u>problem</u> ‘problem’ (loan)	164	308
* <u>cel</u> ‘aim’ (loan)	190-193	266
* <u>komitet</u> ‘committee’ (loan)	236	230
* <u>teren</u> <sup>30</sup> ‘premises’ (loan)	241-243	226
** <u>proces</u> ‘process’ (loan)	282-284	195
** <u>zakres</u> ‘range’	299	190
** <u>wiek</u> <sup>31</sup> ‘age’	312-313	182
* <u>szereg</u> ‘row’ (loan)	364-347	166
** <u>młodzież</u> (f) ‘youth’	359-361	161
** <u>przedstawiciel</u> <sup>32</sup> ‘representative’	373-374	153
** <u>żołnierz</u> ‘soldier’ (loan)	485-490	123
** <u>charakter</u> ‘personality’ (loan)	499-506	120
* <u>numer</u> ‘number’ (loan)	499-506	120
* <u>obywatel</u> ‘citizen’ (loan)	530-540	114
* <u>kolej</u> (f) ‘railway’	546-548	112
** <u>inżynier</u> ‘engineer’ (loan)	614-621	102
** <u>sklep</u> ‘store’	639-649	98
* <u>interes</u> ‘business’ (loan)	705-708	90
** <u>brzeg</u> ‘rim’	736-749	86
* <u>odpowiedź</u> (f) ‘response’	788-799	81

<sup>30</sup> This sequence may alternate in feminine nouns, cf. *sarna* ‘deer (doe)-nom. sg.’ and its. pl. gen. *saren~sarn*, cf. section 4.3.

<sup>31</sup> The potential initial sequence *wk-* [fk] occurs across a morpheme boundary, but not morpheme-initially.

<sup>32</sup> The palatalized [tɕ] is not found before [l], but [ts] does occur in this morphophonemic context, cf. *szpicel*, *szpicla* ‘snooper-nom., gen.’.

szef ‘boss’ (loan)	846-854	76
*premier ‘first minister’ (loan)	855-865	75
**przestrzeń (f) ‘space’	855-865	75
**nauczyciel ‘teacher’	904-913	71
**mecz ‘game’ (loan)	959-981	67
**przebieg ‘distance’	959-981	67
**sukces ‘success’ (loan)	959-981	67
**przyjaciel ‘friend’	982-1002	66

Table XXI contains 38 nouns with the non-alternating final *-CC* cluster found in the same frequency range. I have included nouns whose stems end in *-qC* and *-eC* sequences that historically contained the nasal vowels [ɔ̃] and [ɛ̃], respectively, but in modern Polish have the “split” pronunciation with an oral vowel and a nasal homorganic with the following consonant, i.e. [ɔ̃NC] and [ɛ̃NC]. In this category, the final cluster *Ct*<sup>33</sup> appears as the most frequent: it occurs in 18 words, including 9 instances of [nt], 4 of [kt], 3 of [st], one of [jt] and one of [wt]. The cluster [ɕtɕ] occurs in 7 words (of the feminine gender except for one, which is masculine). There are three instances of the [tr] cluster, two of [nts] and one of each of the following: [ŋk], [ŋkt], [ns], [rtɕ], [sw], [ɕl], [zm] and [lm]. Note that none of the clusters in this category coincides with clusters found in the two former groups of nouns.

Table XXI. Masculine nouns (unless marked otherwise) with the stem-final non-alternating *-CC* cluster, ordered according to their rank in the first 1000 word-list.

miesiąc ‘month’	153-154	330
procent ‘percent’ (loan)	171-172	295
punkt ‘point’ (loan)	179-180	284
przemysł ‘industry’	216-217	247
ciąg ‘sequence’	230-231	237
rząd ‘government’	235	232
fakt ‘fact’ (loan)	330-331	174
myśl (f) ‘thought’	341-343	169
wzrost ‘increase’	341-343	169
metr ‘meter’ (loan)	345	167
projekt ‘project’ (loan)	346-347	166
śmierć (f) ‘death’	396-399	146
prezydent ‘president’ (loan)	417-422	142
kilometr ‘kilometer’ (loan)	427	140
koszt ‘cost’ (loan)	541-545	113
element ‘element’ (loan)	549-557	111

<sup>33</sup> I describe the clusters assuming the “final devoicing” pronunciation and not the phonemic value of the consonant, which reveals itself only word-medially before a suffix, cf. *kąt* ‘corner’ versus *prąd* ‘current’: nom. [kɔ̃nt], instr. [kɔ̃ntem] and nom. [prɔ̃nt], instr. [prɔ̃ntem]. In both cases, the final consonant is pronounced as voiced when the following word starts with a voiced obstruent, cf. *ką[d] domu* ‘corner of the house’, *prą[d] zmienny* ‘alternating current’.

gość ‘guest’	558-564	110
list ‘letter’	565-578	109
film ‘film’ (loan)	602-613	103
ludność (f) ‘population’	602-613	103
sens ‘sense’ (loan)	662-669	95
ksiądz ‘priest’	685-692	92
moment ‘moment’ (loan)	693-704	91
prędkość (f) ‘speed’	718-729	88
przyszłość (f) ‘future’	765-778	83
kąt ‘corner’	788-799	81
student ‘student’ (loan)	800-810	80
produkt ‘product’ (loan)	820-835	78
wiadomość (f) ‘news’	820-835	78
obiekt ‘object’ (loan)	836-845	77
prąd ‘current’	836-845	77
wiatr ‘wind’	836-845	77
front ‘front’ (loan)	866-877	74
kształt ‘shape’ (loan)	940-958	68
jedność (f) ‘unity’	959-981	67
organizm ‘organism’ (loan)	959-981	67
zjazd ‘convention’	959-981	67
długość (f) ‘length’	983-1002	66

The distribution of the three kinds of clusters looks very similar in the second thousand of the most frequent words, presented in Tables XXII-XXIV below.

Out of 30 alternating nouns of Table XXII, [Cek~CkV] constitutes the most numerous type with 12 occurrences ([nek~nkV] – 3 times, [tek~tkV], [dek~tkV] and [wek~wkV] – 2 times each, and one occurrence of [stek~stkV], [sek~skV] and [rek~rkV]). The sequence [Cɛɲ~CɲV] is found in 6 cases (twice as [tʃɛɲ~tʃɲV] and once as [ɕɛɲ~ɕɲV], [p’ɛɲ~pɲV], [tɕɛɲ~tɲV] and [dɕɛɲ~dɲV]), and [Cɛts~CtsV] occurs in 4 cases ([v’ɛts~ftsV] - 3 times and [ɜts~rtsV] once). There are two occurrences of [ster~strV], as well as of [Cew] ([bew~bwV] and [sew~swV]) and [Cɛl~CɪV] ([bel~blV] and [g’ɛl~glV]), and single occurrences of [#sen~#snV] and [#p’ɛs~#psV].

For a large majority of the 23 nouns of Table XXIII with the non-alternating *CeC* sequence, *e*-deletion would create a medial or final cluster not found in that environment (or elsewhere in Polish). Only in three nouns *e*-deletion is potentially possible, cf. later examples in (32).

The non-alternating final *CC* sequences found in 60 nouns of Table XXIV largely coincide with those mentioned earlier for the first thousand range, cf. the summary in (31).

Table XXII. The  $e\sim\emptyset$  alternating nouns of masculine gender (unless marked otherwise) ranked 1003-2009.

świadek ‘witness’	1019-1036	64
pies ‘dog’	1047-1065	62
węgiel ‘coal’	1047-1065	62
uczeń ‘pupil’	1128-1147	57
wrzesień ‘September’	1165-1185	55
diabeł ‘devil’ (loan)	1186-1209	54
poniedziałek ‘Monday’	1186-1209	54
użytek ‘use’	1186-1209	54
surowiec ‘raw material’	1224-1243	52
wtorek ‘Tuesday’	1271-1301	50
poseł ‘MP’	1302-1321	49
czerwiec ‘June’	1380-1413	46
pierwiastek ‘element ( <i>chem</i> )’	1444-1475	44
dziadek ‘grandfather’ (loan)	1541-1570	41
magister ‘MA/MS title’ (loan)	1541-1570	41
piasek ‘sand’	1541-1570	41
ładunek ‘load’ (loan)	1657-1697	38
rysunek ‘drawing’	1657-1697	38
sen ‘dream’	1698-1757	37
fachowiec ‘expert’	1698-1757	37
kwiecień ‘April’	1758-1805	36
dworzec ‘station’	1854-1909	34
kawałek ‘piece’	1854-1909	34
sierpień ‘August’	1854-1909	34
styczeń ‘January’	1854-1909	34
wiceminister ‘vice-minister’ (loan)	1854-1909	34
grudzień ‘December’	1942-2009	32
mebel ‘furniture’ (loan)	1942-2009	32
rachunek ‘bill’ (loan)	1942-2009	32
wydatek ‘expense’	1942-2009	32

Table XXIII. Nouns of the masculine gender (unless marked otherwise) ranked 1003-2009 with the stem-final  $-CeC$ .

*kamień ‘stone’	1003-1018	65
model ‘model’ (loan)	1019-1036	64
*papier ‘paper’ (loan) <sup>34</sup>	1037-1046	63
**oficer ‘officer’ (loan)	1104-1127	58
**sieć (f) ‘net(work)’	1104-1127	58
*kieszka ‘pocket’	1019-1036	54
*promień ‘ray’	1210-1223	53

<sup>34</sup> The alternation is possible for the  $-per$  sequence, cf. *koper*, *kopru* ‘dill-nom., gen.’, but I do not find examples of the exact  $-p’er$  sequence.

*uniwersytet ‘university’ (loan)	1210-1223	53
**budżet ‘budget’ (loan)	1224-1243	52
*piec ‘oven’	1244-1270	51
*prezes ‘president’ (loan)	1244-1270	51
*cień ‘shadow’	1380-1413	46
*śmiech ‘laughter’	1380-1413	46
*śnieg ‘snow’	1476-1505	43
*kongres ‘congress’ (loan)	1506-1540	42
*zabieg ‘treatment’	1506-1540	42
*bieg ‘run’	1657-1697	38
facet ‘guy’ (loan)	1657-1697	38
*wypowiedź (f) ‘statement’	1657-1697	38
*chleb ‘bread’	1758-1805	36
bohater ‘hero’ (loan)	1806-1853	35
*adres ‘address’ (loan)	1910-1941	33
bilet ‘ticket’ (loan)	1854-1909	34

Table XXIV. Masculine nouns (unless marked otherwise) with the stem-final non-alternating –CC cluster, ranked 1003-2009.

grunt ‘ground’ (loan)	1019-1036	64
port ‘port’ (loan)	1019-1036	64
pamięć (f)	1037-1046	63
kontakt ‘contact’ (loan)	1047-1965	62
transport ‘transport’ (loan)	1104-1127	58
socjalizm ‘socialism’ (loan)	1128-1164	57
resort ‘department’ (loan)	1148-1164	56
dokument ‘document’ (loan)	1165-1185	55
sejm ‘Polish parliament’ (an)	1186-1209	54
sprzęt ‘equipment’	1186-1209	54
pociąg ‘train’	1210-1223	53
wielkość (f) ‘size’	1210-1223	53
błąd ‘mistake’	1224-1243	52
deszcz ‘rain’	1244-1270	51
jakość (f) ‘quality’	1244-1270	51
mistrz ‘master’ (loan)	1244-1270	51
efekt ‘effect’ (loan)	1271-1301	50
urząd ‘office’	1271-1301	50
całość (f) ‘whole’	1302-1321	49
kurs ‘course’ (loan)	1322-1355	48
pewność (f) ‘certainty’	1322-1355	48
treść (f) ‘content’	1322-1355	48
wątpliwość (f) ‘doubt’	1322-1355	48
czynność (f) ‘activity’	1356-1379	47
własność (f) ‘property’	1356-1379	47
konkurs ‘competition’ (loan)	1380-1413	46

świadomość (f) ‘conscience’	1380-1413	46
centymetr ‘centimeter’ (loan)	1444-1475	44
eksport ‘export’ (loan)	1444-1475	44
miłość (f) ‘love’	1444-1475	44
szybkość (f) ‘speed’	1444-1475	44
wolność (f) ‘freedom’	1444-1475	44
miliard ‘billion’ (loan)	1476-1505	43
przeszłość ‘past’	1476-1505	43
korzyść (f) ‘profit’	1506-1540	42
remont ‘renovation’ (loan)	1506-1540	42
częstotliwość (f) ‘frequency’	1541-1570	41
obecność (f) ‘presence’	1541-1570	41
radość (f) ‘joy’	1541-1570	41
zdolność (f) ‘ability’	1541-1570	41
ciemność (f) ‘darkness’	1571-1612	40
most ‘bridge’	1571-1612	40
zasięg ‘scope’	1571-1612	40
konflikt ‘conflict’ (loan)	1613-1648	39
łączność (f) ‘connection’	1613-1648	39
przyjaźń (f) ‘friendship’	1613-1648	39
cześć (f) ‘honor’	1657-1697	38
absolwent ‘graduate’ (loan)	1698-1757	37
wyjazd ‘departure’	1698-1757	37
akt ‘act’ (loan)	1758-1805	36
umiejętność (f) ‘skill’	1758-1805	36
pojazd ‘vehicle’	1806-1853	35
tekst ‘text’ (loan)	1806-1853	35
dźwięk ‘sound’	1910-1941	33
piers (f) ‘breast’	1910-1941	33
przyrząd ‘instrument’	1910-1941	33
głęb ‘depth’	1942-2009	32
okoliczność (f) ‘circumstance’	1942-2009	32
park ‘park’ (loan)	1942-2009	32
znajomość ‘acquaintance’	1942-2009	32

The data presented thus far reveal several important facts. First of all, the class of *e~∅* alternating masculine nouns is very large in the highest frequencies’ range of the first two thousand words – it comprises 65 nouns in total. In comparison to the two alternations discussed previously, it outnumbers the nouns of all genders of the stable *o~u* class in the same frequency range (56) and is much more numerous than the recessive *e~a/o* class (28 potentially “alternable” nouns, cf. chapter 2). As I will show momentarily, the *e~∅* alternating masculine nouns are very frequent among lower frequencies, too. Therefore, the alternating pattern can be easily acquired and memorized by speakers, and there is no need for stem analogy.

Two other classes of nouns, containing non-alternating *CeC* or *CC* sequences, are very well represented among the highest ranks, as well (52 occurrences of the former

versus 98 of the latter). But we do observe a strong bias towards particular combinations of consonants in each group, especially with regard to the consonant in  $C_2$  position.<sup>35</sup> The summary of all clusters occurring in the  $e\sim\emptyset$  alternating class in (30) compared to the respective summary of  $CC$  clusters in non-alternating nouns in (31) reveals the differences conspicuously. While  $k$  appears as  $C_2$  in half of all alternating nouns, it is found in the non-alternating group only as a member of an  $NC$  cluster in four words and once in the  $rk$  cluster (total 5%). The consonant  $t$ , which represents the most common  $C_2$  of the latter class (39%), does not occur in that position in the former class at all. The second most frequent  $C_2$  in (30), the consonant  $ts$  (15%), appears in (31) twice, only as a member of an  $NC$  cluster; and the third most frequent  $C_2$  in (30),  $\eta$  (12%), is absent in (31). Likewise, the second most common cluster of (31),  $\epsilon t\epsilon$  with 32% occurrences, is not found in (30) as a possible alternating sequence and  $t\epsilon$  does not occur as a  $C_2$  at all. Out of all other less frequent sequences, only the non-alternating  $tr$  (four occurrences) has a counterpart as the alternating  $ster\sim str$  sequence (three occurrences). In other cases, even if both types of nouns share  $C_2$  ( $l, s$ ), there is a difference in  $C_1$  ( $b, g'$  versus  $\epsilon, p$  versus  $r, s$ ). I list monosyllabic alternating words separately in (30), since they constitute a pattern of their own with an alternating sequence being simultaneously stem-final and word-initial.

(30) The summary of clusters in  $e\sim\emptyset$  alternating masculine nouns ranking 1-2009.

type of cluster	ranks 1-1002	ranks 1003-2009	total ranks 1-2009
Cek~CkV	22	12	34 (52%)
Cets~CtsV	6	4	10 (15%)
Ce $\eta$ ~C $\eta$ V	2	6	8 (12%)
ster~strV	1	2	3 (5%)
Cew~CwV	0	2	2
Cel~ClV	0	2	2
ndel~ndlV	1	0	1
#v <sup>i</sup> ec~#f $\epsilon$ V	1	0	1
#kref~#krfV	1	0	1
#d $\epsilon$ e $\eta$ ~#d $\eta$ V	1	0	1
#sen~#snV	0	1	1
#p <sup>i</sup> es~#psV	0	1	1
total nouns:	35	30	65

<sup>35</sup>  $C_1$  is relevant in the  $CC$  class of nouns, because of the frequent  $NC$  sequences (resulting either from the earlier nasal vowel+consonant combinations or loanwords).

(31) The summary of non-alternating *CC* clusters in masculine nouns ranking 1-2009.

type of cluster	ranks 1-1002	ranks 1003-2009	total ranks 1-2009
Ct	18	20	38 (39%)
çtç	7	24	31 (32%)
tr	3	1	4 (4%)
ŋk	1	3	4 (4%)
rk	0	1	1
nts	2	0	2
ns	1	0	1
rs	0	2	2
rç	0	1	1
rtç	1	0	1
ntç	0	1	1
stj	0	1	1
mp	0	1	1
sw	1	0	1
çl	1	0	1
zm	1	1	2
lm	1	0	1
jm <sup>36</sup>	0	1	1
zɲ	0	1	1
ŋkt	1	0	1
kst	0	1	1
stj	0	1	1
total nouns:	38	60	98

While the consonantal sequences in (30) and (31) are largely complementary with respect to each other, they are complementary as a whole with the consonants of nouns with fixed *CeC* stem ending, presented previously in Tables XX and XXIII. Recall that in most of those cases the *e* vowel breaks up a cluster which is not found in that environment (or generally in Polish). Only in a few cases the two consonants could form a permissible cluster, as summarized and exemplified in (32) below.

---

<sup>36</sup> This final cluster is found in only one noun, which I discuss later in section 4.2.2.

(32) Masculine nouns ranking 1-2009 with non-alternating *CeC* clusters in which *e*-deletion would be potentially permissible.

<i>CeC</i> noun	<i>CeC</i> ~ <i>CCV</i> example	- <i>CC</i> example
<u>szef</u> ‘chief’	<u>szew</u> , <u>szwu</u> ‘seam’	<i>impossible</i>
<u>model</u> ‘model’	<u>pudel</u> , <u>pu<u>đ</u>la</u> ‘poodle’	<i>not found</i>
<u>bohater</u> ‘hero’	<u>sweter</u> , <u>swetra</u> ‘sweater’	<u>teatr</u> ‘theatre’
<u>facet</u> ‘guy’	<u>ocet</u> , <u>oc<u>u</u></u> ‘vinegar’	<i>not found</i>

It should be noted that, although the existence of particular suffixes, such as e.g. the nominalizing/diminutive suffix *-(e)k* in the “alternating” class or the feminine nominalizing suffix *-ość* in the “non-alternating” class, considerably increases the number of words of the respective phonological shape, it does not constitute a *sine qua non* condition of the presence/absence of an alternation. Many of the *-ek* final words included in the lists here are either very vaguely motivated as derived, e.g. *stosunek* ‘relation’ versus *stosować* ‘apply’, *piątek* ‘Friday’ versus *piąty* ‘fifth’, or are unmotivated (underived) at all, e.g. *skutek* ‘result’, *statek* ‘ship’, *rynek* ‘market’, *członek* ‘member’, etc. Even less motivation is observed in the case of less productive suffixes, cf. etymologically related *czerwiec* ‘June’ versus *czerw* ‘maggot’, *kwiecień* versus *kwiat* ‘flower’, *wrzesień* ‘September’ versus *wrzos* ‘heather’, or the etymologically unrelated *palec* ‘finger’ versus *pal* ‘pole’. Likewise, although the majority of *-ctę* final nouns are transparently derived from adjectives, as *dlugość* ‘length’ from *długi* ‘long’ etc., some are non-derived, cf. *gość* ‘guest’, *korzyść* ‘profit’. Therefore, it is rather phonological resemblance and not morphological composition that unites particular sequences into the given alternating/non-alternating categories. Consequently, the phonological composition constitutes, apart from frequency, an additional factor enhancing the salience of a “template” of each class, including the *e*~ $\emptyset$  alternating template. We can explain then, why in some cases the *e*~ $\emptyset$  alternation is not only maintained, but extends onto other lexical items of a similar phonological make-up, the issue to which I will return later in this section.

The *e*~ $\emptyset$  alternating nouns are frequent not only within the most common vocabulary, but throughout the Polish lexicon. Within the lower rank of 8739-10355 in *Słownik* (1990), comprising words of four occurrences in the corpus, there are 44 such nouns. The majority of them have *k* as the stem-final consonant (30=68%), which usually belongs to the productive *-ek* suffix (in two functions: nominalizing or diminutive), e.g. *gwizdek* ‘whistle’, *nagrobek* ‘tombstone’, *odpadek* ‘waste’, *zapisek* ‘note’, *listek* ‘leaf-dim.’ The second most common stem-final consonant in this range is *r*, which appears in six words (14%), including some loanwords, e.g. *puder* ‘powder’, *sweter* ‘sweater’, *kufer* ‘trunk’. The remaining nouns have stems ending in *l* (three cases), e.g. *kabel* ‘cable’; *ś* (two cases), e.g. *żywiec* ‘slaughter animal’; *t* (one case), *poczet* ‘circle, composition’<sup>37</sup>. There are also two monosyllabic alternating nouns, *mech* ‘moss’ and *bez* ‘lilac’.

<sup>37</sup> In most frequent of its uses, this word is a part of fixed phrases (e.g. *przyjąć w poczet członków towarzystwa* ‘to admit as a member of a society’) and does not occur in other declensional cases. But in

Within the same range of lower frequencies, words with the fixed stem-final *CeC* sequence are rarer than for higher frequencies. Out of 21 nouns in this group, most would pose syllabification problems under the hypothetical *e*-deletion, e.g. *portfel* ‘wallet’, *portier* ‘doorman’, *powiew* ‘breath of air’, *sień* ‘hall’, *step* ‘steppe’, *Szwed* ‘Swedish’, *zalew* ‘bay’, but a few would not, e.g. *proceder* ‘dealing’ or *skuter* ‘scooter’.

The nouns with the stem-final fixed *CC* sequence are numerous within this frequency range and amount to 78. Many of them end in the above-mentioned productive (feminine) suffix *-ość*, e.g. *należność* ‘payment’, *rzadkość* ‘rarity’, *wieczność* ‘eternity’, and a few contain the suffix *-izm*, e.g. *nacjonalizm* ‘nationalism’, *radyzm* ‘radicalism’. *NC* final sequences are frequent among native nouns (containing the historical nasal vowel), as e.g. *wdzięk* ‘charm’, *wąs* ‘moustache’, *zamęt* ‘chaos’, as well as among loanwords, e.g. *konsultant* ‘consultant’, *recenzent* ‘reviewer’, *precedens* ‘precedence’. Sporadic occurrences include sequences distinguished earlier among words of higher frequencies, cf. *intelekt* ‘intellect’, *pomost* ‘link (bridge)’, but also many others, cf. *aneks* ‘appendix’, *koktajl* ‘cocktail’, *szewc* ‘shoe-maker’, *wosk* ‘wax’, *żółw* ‘turtle’, etc.

The observation of the nouns of lower frequencies leads to the following conclusion. While the templatic limitations seem to hold within the *e*~ $\emptyset$  alternating nouns even with a greater force than in higher frequencies (e.g. 68% of *C(e)k* to 52%), they are much more relaxed within the non-alternating classes, in which various unique sequences are found. This state of affairs is well understandable, since an alternation of a rarer occurrence word is more difficult to maintain and it has to be more template-driven. No such constraint is imposed on non-alternating nouns, which can thus accommodate a number of low frequency words of “foreign” or “learned” lexicon.

Returning to the problem of the *e*~ $\emptyset$  alternating pattern extension, we can hypothesize that it will take place in cases of well-established templates, for example, when the stem-final consonant is *k*. Indeed, such cases are found among adapted loanwords, as e.g. the above-mentioned adaptation of the German suffix *-ng*, realized in Polish in the alternating form as *nek*~*ŋkV*. But the issue is a little more complex than that. Let us observe that stem-final non-alternating *ŋk* sequence is also found in Polish, as for example in many native words historically containing the nasal vowel followed by *k* (e.g. *pałk* ‘bud’). This is the reason why not all *ŋk* stem-final loanwords undergo the *e*~ $\emptyset$  alternating pattern extension, e.g. *bank* ‘bank’ preserves the original *CC* sequence essentially in accordance with phonotactic patterns of Polish. Similar choices exist for many other consonant combinations. For example, some *rk* stem-final words have been adjusted to the prevailing alternating template, e.g. *korek* ‘cork’, some have not, e.g. *park* ‘park’ (cf. native *bark* ‘shoulder’). It is very hard to determine for certain, why in each particular case the adaptation does or does not take place, since many various factors seem to be involved. As a brief illustration, I will consider a handful of stems ending in *t* as the first consonant in a sequence and *r* as the second consonant, which are distributed throughout all three classes, i.e. the alternating *ter*~*trV*, and the fixed *tr* and *ter*.

Frequency of use certainly explains why the common noun *sweter* ‘sweater’<sup>38</sup> accommodates the alternating pattern (cf. gen. *swetra*), and the phonologically similar to

---

some more specific meanings it may, e.g. *poczet* (nom.) / *pocztu* (gen.) *sztandarowy/sztandarowego* ‘color party’.

<sup>38</sup> The word has a wider scope of use in Polish than indicated by its English translation, since it may be used as a hyponym (or a synonym) of all kinds of warm, long-sleeved garments, including neck-ties,

it rare-occurrence noun *seter* ‘setter (dog breed)’ preserves the fixed sequence (gen. *setera*). But the situation is not so clear if we compare the alternating loanword *kuter* ‘cutter’ (gen. *kutra*) to *skuter* ‘scooter’ (gen. *skutera*). What seems to differentiate these two nouns is the time of borrowing; the former dates back to the 19<sup>th</sup> c., while the latter is very recent. It appears that nativization of a loanword, like any other change, is gradual and needs some time. At the initial stage, foreign lexicon infiltrates the target language in the “code-switching” manner, preserving as much of the original structure as possible. Only after the word becomes relatively familiar can structural changes begin to take place. Therefore, in the case of very recent borrowings, we would rather expect the word to maximally resemble the source, even if it is very frequent. The word *komputer* ‘computer’ provides a relevant example. It started to be used on a larger scale in the 1980s when PCs first appeared in Poland, initially in public places such as universities etc. rather than in people’s homes. In recent years, the personal possession, as well as the use of the object, is as frequent as everywhere else in modern societies, and so is the use of the word by all generations, including (or perhaps, especially) by children. I can still remember that back in the 1980s the word was pronounced by some persons as [kɔmp<sup>h</sup>uter], but the variant [kɔmputer], better adjusted to Polish phonotactics, was quickly gaining popularity, to the complete elimination of the former option. Until now the noun keeps the non-alternating form in declension, cf. gen. *komputera*. Should we expect alternating pattern analogy to affect this word in the future? I believe, it is very likely, provided computers are still popular and are called by the same name. Occasionally, children and young people can be heard using the genitive *komputra*, so far in a joking manner, but perhaps they are about to initiate a widespread change.

Finally, some loanwords preserve more “foreign” pronunciation for prestige. Should we expect the non-alternating word *teatr* ‘(performance) theatre’ to undergo a change into an alternating *\*teater*, similarly to the mentioned earlier *korek*? Rather not, because it would lose its French flavor, so adequate for a “cultured” word like that. The foreignness of the stem-final *-tr* sequence is felt, even though a few native (or highly nativized early loanwords) end in such a cluster, cf. *wiatr* ‘wind’, *Piotr* ‘Peter’. In some other *-CC* final stems, the combination never occurs in the native and common lexicon, but only in sophisticated loanwords, by which it is even more “unique”, cf. *konstabl*, *ensamble* with the final *-bl* cluster versus alternating “common” words, also of foreign origin *kabel* ‘cable’ or *jubel* ‘party- coll.’

To conclude this section, the *e-∅* alternating nouns of the masculine gender present a clear templatic model among all frequencies due to their type abundance, as well as relative phonological distinctiveness with respect to other competing non-alternating classes, which results not only in the preservation of the alternation, but also in its slight expansion onto new lexical items. Fairly balanced frequency of both stem variants within the paradigm, i.e. the minor *CeC* variant of nominative

---

pullovers, woolen jackets etc. In the Polish climate, *sweter* is worn all year long, which makes it a very frequent word in every day usage (not as much reflected by *Slownik* (1990) and the PWN Corpus data). The high frequency triggers the colloquial shortening of the nom.=acc. case to *swetr* (in parallel to *wiatr* ‘wind’). Although the colloquial form eliminates alternation within the paradigm (but in dim. *sweterek*, likewise *wiaterek* ‘wind-dim.’), in my opinion it is conditioned not by analogy, but Zipf’s frequency laws (cf. chapter 6), with leveling being only a ‘side-effect’.

(nominative=accusative) and the major *CC* variant of other declensional cases, constitutes an additional factor in favor of the alternation.

#### 4.2.2. Analogy in nouns of the infrequent type

In all nouns discussed in the previous sections, the “ghost” vowel was located at the end of the stem. This is indeed the only possible situation in the contemporary language – I will show in this section that it naturally follows from the historical development of these vowels. At the same time, let us observe that this restricted position naturally facilitates the reinterpretation of the earlier phonological alternation as a new morphophonemic alternation, limited to “the end of the word” contexts.

In the time of the yers’ existence, there were no restrictions as to their location within a root or stem. Some yers occurred before a full vowel in the following syllable. Recall from section 4.1. of this chapter, that a yer in this position was a “weak” one, prone to disappearance. The regular loss of the yer in such an environment resulted in a unified *CCVC* stem-form throughout the declension, before a zero suffix, as well as before a full vowel suffix, as illustrated below.

(33) The regular development of a yer-stem

old stem: / рѣtak/		modern stem: /ptak/
*рѣtakъ	>	ptak ‘bird-nom.’
*рѣtaka	>	ptaka ‘bird-gen.’

There also existed nouns whose stems contained two consecutive yers divided by a consonant. They surfaced as *CCeC* before a zero inflection and as *CeCC* before a full declensional suffix, which resulted in stem alternation within their paradigms, as illustrated in example (34) below, repeated from (24).

(34) The regular development of a two-yer-stem

Early Polish		Old Polish		contemporary form+gloss
ѡвѣтсѣ	>	ѡv <sup>j</sup> ets		jefts (ort. szewc) ‘shoe-maker-nom.’
ѡвѣтса	>	jeftsa		jeftsa (ort. szewca) ‘shoe-maker-gen.’

Let us observe that the alternation found in Old Polish inherited from the two-yer-sequence stems differs in type from the *VCeC~VCC* alternation discussed in the previous sections. In terms of stem correspondence, the former type induces more violations than the latter (two violations of *CC* contiguity versus one), hence can be argued to be more susceptible to leveling. But this fact would not probably be relevant if there existed a large, salient class of such nouns. In fact, they were very few and all underwent analogy. Let us consider three masculine nouns of that shape, shown in (35), all of which have alternating forms attested historically. In the first example, in (35a), the leveling was

towards the *CeCC* stem variant, which occurred in all declensional cases except for the nominative=accusative. Likewise in (35b), with the only difference that this noun is animate and had the minor *CeCC* variant only in the nominative. In (35c), however, the leveling was oriented towards the *CeCC* stem pattern of the nominative=accusative. In order to see whether frequency of particular stem patterns could play a role in leveling, I have included the PWN Corpus figures indicating (joint) occurrences of word-forms of the two types, i.e. before a zero suffix (earlier stem: *CCeC*) and before a full suffix (earlier: stem *CeCC*). The form which became the Base for leveling is underlined, and an asterisk precedes the form which was eliminated.

(35) PWN frequencies of leveled two-*yer*-nouns

	noun+gloss	zero suffix ( <i>CCeC</i> )	full suffix ( <i>CeCC</i> )
a/	sejm ‘Polish parliament’ <sup>39</sup>	*sjem 3282 (38,8%)	<u>sejm</u> 5168 (61,2%)
b/	szewc ‘shoe-maker’	*ʃv <sup>l</sup> ets 79 (33,6%)	<u>ʃefts</u> 156 (66,4%)
c/	szmer ‘rustle’	<u>ʃmer</u> 107 (48,9%)	*ʃemr 112 (51,1%)

Even though the PWN Corpus data reflect the present day frequency of these lexemes, they shed some light on what could be the usage of the nouns at the time the leveling took place. In (35a), the relatively high frequency of the nominative=accusative case is due to common expressions such as, e.g. (nom.) *sejm zebrał się / obraduje / zdecydował* ‘the parliament gathered / is sitting / decided’, or (acc.) *zwołać sejm* ‘to call the parliament’. But the joint uses of other cases are significantly higher, which is mostly due to the high frequency of the locative, as in e.g. *dyskusja w sejmie* ‘discussion in the parliament’, as well as the genitive, as in e.g. *obradę sejm* ‘session of the parliament’. There is no reason to suppose that in the past these proportions were different. We may then conclude that it is very likely that frequency was the reason for this particular direction of analogy, even though it created an unprecedented final coda, namely, *-jm*.

The most interesting example is that of (35b). The present day use of the lexeme is very low, because the shoemaking profession almost completely disappeared from the contemporary life. Of course, in the not so distant past, everyone bought and had their shoes repaired at the shoemaker’s. Without any doubt, the use of locative expressions with the genitive case (of the *CVCC* stem shape) was very high, cf. *iść (zanieść / oddać buty) do szewca* ‘go (take the shoes) to the shoe-maker’s’, *zamówić (zrobić / zreperować) buty u szewca* ‘order (have made / have repaired) shoes at the shoemaker’s’, *przynieść buty od szewca* ‘bring the shoes from the shoemaker’s’, etc. And we can confidently hypothesize that the occurrence of such phrases was much higher than those in which ‘shoe-maker’ appeared in the nominative case as a subject (with the *CCVC* stem), as in e.g. *ten szewc jest miły (dobry / drogi)* ‘this shoe-maker is nice (good / expensive)’. The corpus data still reflect the higher use of the inflectional cases, which historically had the *ʃefts* stem. Let us observe that this form was generalized in spite of the fact that it created an unusual word-final cluster *-fts*, otherwise not found in Polish.

<sup>39</sup> More exactly, *sejm* is the lower chamber of the Polish parliament.

The last noun in (35c) has low frequency and almost equal proportion of the historical \*CCVC and \*CVCC allomorphs. The leveling was directed towards the nominative *fmer* variant, perhaps by the strategy of “nominative-unmarkedness in cases undetermined by frequency”, suggested previously in section 3.3 of the previous chapter. It is noticeable that the leveling process was not affected by the *femr* stem variant occurring in the verb *szemrać* ‘rustle’.<sup>40</sup>

In some cases, analogical changes were preceded by idiosyncratic cluster simplification, as in the two following examples.

The Old Polish form of the nominative of ‘pepper’ in (36), similarly to other two-*y*er stems, is regularly derived except that the initial cluster is degeminated. The modern, leveled form adopted the stem from other non-nominative cases (as in *sejm* and *szewc*). Frequency could play a role here, cf. frequent expressions with cases, such as e.g. instr.: (*danie*) *z pieprzem* ‘(a dish) with pepper’ or gen.: (*dodać*) *pieprzu* ‘(add) pepper’, etc.<sup>41</sup>

(36) The development of *pieprz* ‘pepper’

Early Polish	Old Polish	contemporary form+gloss
ръррьгь	(p)pʲeʃ	pʲepʃ (ort. <i>pieprz</i> ) ‘pepper-nom.’
ръррьг’у	pʲepʃu	pʲepʃu (ort. <i>pieprzu</i> ) ‘pepper-gen.’

In the case of *deszcz* ‘rain’ in (37), both Old Polish forms are regular with respect to the reflexes of *yers*, although the clusters are simplified through assimilation. In modern Polish, the nominative=accusative case became the Base for analogy, presumably because of the high frequency of the common expression *pada deszcz* ‘it’s raining’ – the most “unmarked” verbalization of the meteorological fact. In language usage, this particular expression, in which the verb *pada* lit. ‘falls’ precedes the noun, most typically occurs before a pause. In more elaborated contexts, usually the reverse order appears, cf. *deszcz pada i pada* ‘it is raining and raining’ (lit. ‘the rain is falling and falling’) or *deszcz pada cały dzień* ‘it is raining all day’. Let us observe that in both variants of the word order the noun *deszcz* occurs in the phonological environment entailing the devoicing of the coda consonants, either pre-pausally, or due to the assimilation to the following word *pada* starting with a voiceless consonant. The voiced variant of the coda is practically limited to the rare contexts, in which a modifier starting with a voiced consonant follows the subject or object *deszcz*, e.g. *deszcz wiosenny* [deʒɕ v’osenni] ‘spring rain’. It is very likely that the disproportionately frequent and fairly fixed context of the occurrence of *deszcz* before a pause/voiceless stop caused the generalization of the devoiced stem [deʃtʃ] throughout the paradigm. This is a very atypical situation in Polish, since normally analogical leveling resists final devoicing and voicing assimilations preserving

<sup>40</sup> Laskowski (1975:38, fn. 25) lists *szemr* among other -CC final nouns (which excludes a typographic mistake). I myself have never come across such a variant and it does not figure in the dictionaries. If it is a possible (dialectal?) form, it supports my earlier claim that variation is expected when frequency criteria are indeterminable.

<sup>41</sup> PWN frequencies are „untestable” in the case of this word.

the phonemic value of consonants.<sup>42</sup> Hence, the expected analogical form of the genitive should be \**dezdžu* [deʒdʒu], and not actually used *deszczu* [deʃtʃu] (likewise in other inflectional cases).

(37) The development of *deszcz* ‘rain’

Early Polish	Old Polish	contemporary form+gloss
дѣздѣ	deʒdʒ (deʃtʃ)	<u>deʃtʃ</u> (ort. <i>deszcz</i> ) ‘rain-nom.’
дѣздѣу	dʒdʒu	deʃtʃu (ort. <i>deszczu</i> ) ‘rain-gen.’

According to the PWN Corpus data, *deszcz* appears as the most frequent of all word-forms, with 978 occurrences constituting 44,15% of total 2215 occurrences of the lexeme DESZCZ. It should be borne in mind, however, that a great majority of the corpus data come from written sources, often literary works, in which the meteorological fact of raining is usually expressed in a more sophisticated way than by means of the common expression mentioned earlier occurring in everyday usage. For example, the genitive=locative word-form *deszczu* appeared in the corpus as often as 767 times (34,63%), typically in expressions as e.g. *krople deszczu (bębniły o szyby)* ‘drops of rain (were drumming against the window panes)’ or *iść w deszczu* ‘walk in the rain’, significantly contributing to the total occurrence of this inflectional case, which is presumably much rarer in the spoken language.

Finally, there is one more unusual fact concerning the development of the Polish word for ‘rain’, namely, the survival of the old genitive *dźdžu* [dʒdʒu] in sporadic expressions (cf. *krople dźdžu* ‘drops of rain’ as a more “poetic” variant), including one proverb (total 13 occurrences in the PWN Corpus). Although the meaning of *dźdžu* is just as that of *deszczu*, the speakers of Polish do not seem to realize that these two words are related.<sup>43</sup>

The old Germanic loanword in (38) has undergone anticipatory voicing assimilation in the nominative=accusative singular, which, together with different vocalism, made it look very distinct from other cases. The leveling took place in the nominative on the basis of other declensional cases, resulting in the uniformed stem *tsebr* throughout the paradigm in Old Polish. High frequency of this stem form (occurring in all cases other than the singular nominative=accusative) was likely the trigger of leveling, especially since the word is used in the common expression *leje jak z cebra* ‘it rains cats

<sup>42</sup> Another exceptional example of a similar kind is the variation *lepek~lebek* ‘(animal) head-dim.’ The former variant “copies” the final devoicing, while the latter is faithful to the phonemic value of the consonant, cf. *lba* [wba] ‘(animal) head-gen.’ Occasionally, analogy “overapplies” voicing assimilation, e.g. \**pazno[g]ieć* > *pazno[k]ieć* ‘nail-nom.’ under the influence of other cases, as *pazno[k]cia* ‘nail-gen.’, or \**szczaq[d]ek* > *szczaq[t]ek* ‘remains-nom.’ triggered by *szczaq[t]ki* ‘remains-pl. nom.’ (cf. Ułaszyn 1956:13n.).

<sup>43</sup> I came to this conclusion after being once asked in a class by a group of students (all native speakers of Polish): “What is the nominative of that funny word *dźdžu*?”, while they were trying to produce something as the impossible *dźdź*. My explanation that the word has the same source as *deszczu* appeared quite shocking to the students, presumably because of the heavy functional load of voicing in onset obstruents which makes the two words look very distinct.

and dogs' lit. 'it pours as from the bucket'. Although at present the usage of this word is rather limited, the PWN corpus data to some extent confirm the low frequency of the nominative (no occurrences versus 46 occurrences of other cases, NB. not only in the above-mentioned expression). Shortly after the leveling, the nominative singular was remodeled as *tseber*, according to the *e~∅* alternating pattern. Bańkowski (2000, v. 1:109) dates the nominative form *tsebr* for the first half of the 16<sup>th</sup> c., and *tseber* already for the second half of the 16<sup>th</sup> c. It is remarkable how pattern analogy destroyed the order just made by stem analogy, "reviving" the old "ghost" vowel.

(38) The development of *ceber* 'wooden pail'

Early Polish	Old Polish	contemporary form+gloss
цѣбѣръ	ѣber > tsebr	tseber (ort. ceber) 'wooden pail-nom.'
цѣбѣра	tsebra	<u>tsebra</u> (ort. cebra) 'wooden pail-gen.'

#### 4.2.3. Analogy in masculine diminutives

The diminutive suffix *-(e)k* originally contained a yer vowel which was subject to the same processes of vocalization and deletion as other yers. In masculine nouns, it surfaced as *e* in the singular nominative (nom.=acc. of inanimates) and was deleted before vowel-initial suffixes of all other cases in the paradigm. If the final vowel of the stem was a full vowel, the stem remained constant in all declensional cases and the only alternation was that introduced by the suffix itself, as illustrated in (39) below (with morpheme boundaries marked for convenience). Such alternations remained until present and are synchronically no different from other *e~∅* alternations discussed previously. Note that in (39b) the stem-final velar is palatalized in the presence of the diminutive suffix. I leave aside this quite regular process.

(39)

	Early Polish		Old Polish=contemporary form+gloss
a/	дѡм-ѣк-ѣ	>	дѡм-ek (ort. domek) 'house-dim. nom.'
	дѡм-ѣк-у	>	дѡм-k-у (ort. domku) 'house-dim. gen.'
b/	рѣтак-ѣк-ѣ	>	ptaʃ-ek (ort. ptaszek) 'bird-dim. nom.'
	рѣтак-ѣк-а	>	ptaʃ-k-a (ort. ptaszka) 'bird-dim. nom.'

If the final vowel of the stem was a yer, the resulting diminutive form resembled other cases of nouns with two consecutive yers, discussed in the previous section 4.2.2. The predicted stem alternation of the diminutive is attested in Old Polish for the word given in (40), repeated from (24).

(40) The regular development of a yer-stem diminutive

Early Polish		Old Polish	contemporary form+gloss
рѣс-ѣк-ѣ	>	<u>ps</u> -ek	p <sup>i</sup> es-ek (ort. piesek) ‘dog-dim.-nom.’
рѣс-ѣк-а	>	<u>p<sup>i</sup>es</u> -k-a	p <sup>i</sup> es-k-a (ort. pieska) ‘dog-dim.-gen.’

Bańkowski (2000, v. 1:109) cites a diminutive of the double-yer stem noun *ceber* ‘wooden pail’ discussed above in section 4.2.2, which in Old Polish had a predicted alternating stem, resulting from the regular rhythmic processes of yer deletion/vocalization. As shown in (41), voicing assimilation in the onset takes place as it was previously shown for the base noun. In the contemporary language, the *ek*-suffixed diminutive is not used (hypothetical \**ceberek/ceberka*), replaced with a newly formed *yk*-suffixed *cebrzyk/cebrzyka*.

(41) The diminutive of *ceber* ‘wooden pail’

Early Polish	Old Polish	gloss
цѣбѣгѣкѣ	tsebrek	‘wooden pail-dim.-nom.’
цѣбѣгѣка	ɕberka	‘wooden pail-dim.-gen.’

Even though examples of this kind are rarely cited in historical sources,<sup>44</sup> we can quite confidently hypothesize that other diminutives formed from yer-final stems were characterized by similar regular alternations in the remote past. Many of them were presumably rarer than the common words exemplified above, frequently used in the diminutive, and they might have undergone stem leveling earlier than these particular words which preserved the alternation until Old Polish. In the contemporary language, all nouns with *e~∅* alternating stems have a unified derived diminutive stem throughout the paradigm, similarly to the earlier discussed *o~u* alternating nouns. The *CC*-final inner stem variant is generalized in extremely rare exceptions. One such case of the masculine gender is *garnczyk* [garntʃek] ‘pot-dim.nom.=acc.’ (instead of the expected \**garneczek*), which surfaces with four medial consonants in other cases, cf. gen. *garniczka* [garntʃka]. (A simplified form without the medial *n* occurs as a free variant, cf. *garczek*, *garczka*.) Otherwise, the generalized variant of the inner stem is always *CeC*-final. I will argue that this choice does not reflect the frequency of that particular stem variant, but is triggered by syllable structure criteria combined with the *e~∅* alternating template of diminutives. My line of argumentation is based on the assumption that stem leveling is desired within the class of diminutives – I will return to this issue momentarily. Therefore, I will concentrate only on such theoretical possibilities in which the inner stem is leveled either

<sup>44</sup> The rather few masculine diminutives found in the dictionary of 16<sup>th</sup> Polish (*Słownik 1968-2004*) usually have analogized, modern forms. An old alternation can be only sporadically spotted, e.g. *bochnek* (gen. *bochenka*) given as a variant of *bochenek* ‘loaf of bread’, but it is not certain if this noun of the unknown etymology contained a diminutive suffix (or whether the present augmentative form *bochen* is derived by back-formation, cf. Bańkowski 2000).

as *CeC* or *CC*. Each of these two options can theoretically combine with the diminutive suffix in one of three ways: the suffix either has a fixed form *k* or *ek*, or it alternates as *k~ek*. I will show, that out of the total of these six combinations, only the one actually attested, with the fixed inner stem ending in *CeC* and the alternating diminutive suffix, complies with the general facts of Polish.

Let us first recall the issue already pointed out in chapter 3, namely, the rarity of diminutives among high frequency nouns. It is striking that among the total number of 65 nouns ending in the *-(e)k* sequence found within the range of 1-2009 ranks, only one is an etymological diminutive. It is *dziadek* ‘grand-father’, presently constituting an unmarked noun, while its base form *dziad* is rarely used in the same meaning and more often in the secondary, pejorative sense of ‘old man; beggar, pauper’. In the case of three other nouns, namely, *piasek* ‘sand’, *kawałek* ‘piece’ and *członek* ‘member (in pl. also ‘limbs’), the suffixless nouns exist, but are back-derived, which is clearly felt in the case of *piach* and *kawał*, being augmentatives of the first two nouns respectively, while *człon* has the meaning ‘segment, constituent’, unrelated to the meaning of its morphological base. But even by the most generous count, i.e. including all these four nouns having suffixless counterparts, a category of *-(e)k* diminutives is very poorly represented among high frequency nouns. Let us in addition observe, that unlike the above-mentioned *p'es~ps* ‘dog’, none of these particular diminutives is based on a stem of the alternating *e~∅* type. This means that, if the stem alternation exemplified by the Old Polish case *psek~p'eska* ‘dog-dim. nom.~gen.’ were to remain in the contemporary language, it would be very difficult to maintain without a pattern to follow among the high frequency words.

In lower frequencies, diminutives are much more frequent. Among the total of 44 nouns with the *-(e)k* ending occurring between ranks 8739-10355 in *Słownik* (1990), there are 13 true or slightly lexicalized diminutives, e.g. *baranek* ‘ram-dim.’, *człowieczek* ‘man-dim.’, *daszek* ‘roof-dim.; visor’, *listek* ‘leaf-dim.’ By a generous count, we can add to them two more nouns for which a suffixless noun exists, viz. *chlorek* ‘chloride’ (cf. *chlor* ‘chlorine’) and *tyłek* ‘behind’ (cf. *tył* ‘rear’), which makes the total of 15. However, even in this much larger group, there are no occurrences of diminutives having the required *e~∅* alternating inner stem, either. The conclusion is then that such diminutives are truly rare. In fact, diminutives formed from bases of this particular stem form are fewer than those based on other stems, which is partly due to phonological reasons and partly to semantic reasons. For example, a large proportion of these nouns is derived from verbs by means of the nominalizing *-(e)k* suffix and often have abstract meanings, which generally excludes diminutivization. Nouns whose stems end in *ts* and *n*, which are also very frequent in the *e~∅* alternating class (cf. the earlier summary in (30)), do not take the diminutive *-(e)k* for phonological reasons.

To get a rough idea of how such restrictions reduce the class of potential diminutives suffixed with *-(e)k*, let us see how many of them can be formed from all 65 *e~∅* alternating nouns ranking 1-2009. According to my subjective judgment, only 16 (24,6%) are acceptable and have a chance of actually occurring in language usage. I have included in this count very typical diminutives, as e.g. *piesek* ‘dog-dim.’, *diabełek* ‘devil-dim.’, *węgiełek* ‘small piece of charcoal’, *handelek* ‘trade-dim. (e.g. illegal)’, as well as some which are practically said to children only, as e.g. *stateczek* ‘ship-dim.’, *piaseczek* ‘sand-dim.’, *rysunczek* ‘drawing-dim.’, *środeczek* ‘center-dim.’ (But I did not include diminutives which can be potentially formed from *k* final stems and which seem very odd

and unlikely to be found in a real language situation, as e.g. *?związeczek* ‘union-dim.’, *?waruneczek* ‘condition-dim.’, *?stosuneczek* ‘relation-dim.’, *?członeczek* ‘member-dim.’ etc.) For comparison, almost half of the nouns of the *o~u* alternating class of the same frequency range, can constitute the base of *-(e)k* diminutives, as earlier discussed in chapter 3.

The rarity of diminutives based on an *e~∅* alternating inner stem had to result in leveling. Since pattern analogy is not available in the case of the pre-final vowel of the word (cf. a small group of nouns discussed in the previous section all of which underwent leveling), stem analogy remains the only option. Moreover, since diminutives of this particular make-up are so exceptionally rare, a Base from inside their paradigm would not be easily retrievable even for a small class of pattern-setting lexical items. Let us recall from chapter 3, that a similarly infrequent class of double diminutives derived from *o~u* alternating stems could not sustain a within-paradigm Base, either, and mapped the stem from slightly more salient single-suffixed forms. If a similar strategy applied in the case at hand, it would mean adopting a more frequent stem form from the base noun paradigm. There is, however, a major difference between these two cases with respect to the overall phonological constraints. While the choice of the *o* and *u* stem variants in chapter 3 differs only with respect to the non-transparent constraint on *o/u* distribution in open/closed syllables, the choice of the *CC/CeC* stem variant here has important consequences for the syllable structure – a much more serious gain or penalty. Recall also from the discussion in chapter 3 that in the case of masculine paradigms, the frequency of each of the stem allomorphs varies a lot among particular nouns and often none of the variants is significantly more frequent than the other (unlike in feminine and neuter nouns, which clearly have a minor stem variant). This problem caused some variation, as well as a lexical split in the choice of the Base in the case of single-suffixed diminutives discussed in section 3.3 of the previous chapter. Perhaps in the initial stage, diminutives of the *e~∅* stems followed the same route, with the result that some surfaced with the *CC* inner stem and some with *CeC*. The exception of *garnczek-nom.*, *garnczka-gen.* ‘pot-dim.’ mentioned earlier could be then seen as a relic of a possibly larger group of nouns following the *CC*-pattern. Under such a scenario, the likely competition between two variants to become a generalized template of the diminutive was slowly won by the one which eliminates a difficult medial sequence of three or even four consonants. The comparison in (42) between hypothetical *CC*-leveled and actual *CeC*-leveled diminutives of some common nouns mentioned earlier demonstrates that difficult clusters appearing in the former are easily resolved in the latter.

(42)

	hypothetical <i>CC</i> -leveled	actual <i>CeC</i> -leveled	gloss (nom., gen.)
a/	psək, <u>pska</u>	p <sup>j</sup> esək, p <sup>j</sup> eska	‘dog-dim.’
b/	stattfək, <u>stattfka</u>	statetfək, statetfka	‘ship-dim.’
c/	d <sup>j</sup> abwək, d <sup>j</sup> abw <u>ka</u>	d <sup>j</sup> abewək, d <sup>j</sup> abew <u>ka</u>	‘devil-dim.’
d/	handlek, <u>handlku</u>	handelek, <u>handelku</u>	‘trade-dim.’

The templatic [...*CeC*] structure of the diminutive stem is further supported by the existence of infrequent exceptions, which do not show a “ghost” vowel in the base noun,

but inexplicably acquire it in the diminutive, as (repeated from earlier (5b)) *wiatr* ‘wind’, *wiaterek* ‘wind-dim.’ (contrasted with: *Piotr* ‘Peter’ and its diminutive *Piotrek*). Since this issue is pertinent to feminine and neutral nouns as well, it will be treated more thoroughly in section 4.3.3.

Intuitively, the pressure towards better syllable structure in a language such as Polish may not be too strong and complex clusters resulting from *CC*-leveling could be in principle preserved (or, in some cases idiosyncratically simplified as in the previously mentioned *garneczka* > *garczka*), if there were any other advantage to gain. In the case at hand, both alternatives eliminate the alternation equally well and there seems to be no profit from *CC*-leveling. In fact, there is an additional phonotactic benefit from leveling of the *CeC* inner stem, namely, the size of the diminutive. Under *CC*-leveling, diminutives are as short as their respective base nouns, which for common words as those in (42) usually means one, as (42a), or two syllables, as in (42b-d). But diminutives are not high frequency words, on the contrary, their usage is typically many times rarer than their respective bases (cf. chapter 3). This creates a violation of “proportionality” constraints, or, “Zipf’s frequency laws”, according to which less frequent words should be longer than high frequency words (cf. chapter 6 for discussion). The disproportion between the frequency and the size of the word does not take place under the *CeC*-leveling: the diminutives are always exactly one syllable longer than their respective base nouns, which adequately correlates with their lower frequencies.

Let us observe that the syllabicity problem created under *CC*-leveling for all declensional cases other than the nominative (or, nominative=accusative), could also be resolved, if the diminutive maintained the constant form of *ek* throughout the paradigm. Apparently, this could be seen as an additional “benefit”: the alternation would be eliminated not only within the stem, but also in the suffix. We would then have hypothetical forms as *\*psek*, *\*pseka* ‘dog-dim.’, *\*statfek*, *\*statfeka* ‘ship-dim.’, etc. But under this option, the diminutive of the *e~∅* alternating nouns would have a different realization than all other, much more numerous diminutives, and given the earlier mentioned smallness of its class, such a distinct pattern would not be retrievable. For the same reason, the non-alternating suffix *ek* is not possible in the case of *CeC*-leveled forms, either. Naturally, the non-alternating diminutive suffix *k* is excluded for still the same reasons, as well as because of additional syllable structure problems, especially in the nominative singular, cf. hypothetical *\*piesk*, *\*statet{k}*, or even worse, *\*psk*, *\*statfk*. To conclude, any change of the form of the diminutive suffix hypothetically found in this particular class of diminutives requires a transparent pattern which would have to be found in a more salient, larger class, and which does not exist.

Double diminutives, similarly to those of the *o~u* alternating class, have the same inner stem as single-suffixed diminutives. As already argued in section 3.4 of the previous chapter, the simplest analysis of double (or multi-suffixed) diminutives is by means of infixation of the invariant (i.e. non-alternating) *-etf-* suffix directly before the diminutive suffix, by which they belong to the same class of diminutives (with all consequences), except that they have “extended” diminutive stems, as schematized for *pieseczek* ‘dog-dim.-dim.’ in (43) below. Theoretical aspects of such an analysis will be treated in more detail in section 4.3.3.

(43)

[[[[p <sup>i</sup> es] <sub>stem-etʃ</sub> ] <sub>extended stem-ek</sub> ] <sub>diminutive stem-∅</sub> ] <sub>word</sub>	pieseczek (nom. sg.)
[[[[p <sup>i</sup> es] <sub>stem-etʃ</sub> ] <sub>extended stem-k</sub> ] <sub>diminutive stem-a</sub> ] <sub>word</sub>	pieseczka (gen. sg.)

### 4.3. The *e~∅* alternation in feminine and neuter nouns

#### 4.3.1. The distribution of alternating nouns within the lexicon

Within the most frequent lexicon, *e~∅* alternating nouns of feminine and neuter genders are rarer than those of the masculine gender. Only seven such nouns appear in the first thousand of the most frequent words<sup>45</sup> of *Słownik* (1990) and 18 in the second thousand, as shown in Tables XXV and XXVI. Let us recall that the only case within the paradigm with the *CeC* stem ending is that of the genitive plural – all other word-forms in the singular and plural have the *CC* form of the stem, e.g. *matka*, *matki* ‘mother-nom., gen.’ etc. versus *matek* ‘mothers-gen.’, *światło*, *światła* ‘light-nom., gen.’ etc. versus *światел* ‘lights-gen.’

Table XXV. The *e~∅* alternating nouns of feminine and neuter genders, ordered according to their rank, as they appear in the first 1000-word list.

wojna (f.) ‘war’	226	240
matka (f.) ‘mother’	355-358	162
światło (n.) ‘light’	364-366	159
książka (f.) ‘book’	392-395	147
okno (n.) ‘window’	473-476	126
źródło (n.) ‘source’	730-735	87
kółko (n.) ‘circle-dim.’	846-854	76

Table XXVI. The *e~∅* nouns of feminine and neuter genders ranked 1003-2009.

komórka (f.) ‘cell’	1037-1046	63
łóżko (n.) ‘bed’	1037-1046	63
placówka (f.) ‘agency’	1037-1046	63
córka (f.) ‘daughter’	1224-1243	52
gra (f.) ‘game’	1224-1243	52
skrzydło (n.) ‘wing’	1322-1355	48
panna (f.) ‘maiden’	1356-1379	47
miasteczko (n.) ‘town-dim.’	1541-1570	41
wódka (f.) ‘vodka’	1571-1612	40
dno (n.) ‘bottom’	1657-1697	38
setka (f.) ‘a hundred’	1657-1697	38
butelka (f.) ‘bottle’	1698-1757	37

<sup>45</sup> I did not include the word *gospodarka* ‘economy’ which also appeared in this group, because it rarely takes the plural form. The same word has also a less frequent meaning ‘farm’ and then it takes the plural gen. *gospodarek*, but it would not appear in that meaning among these frequencies, hence it was best not to include it in the count.

dziesiątka (f) ‘ten’	1698-1757	37
piosenka (f) ‘song’	1698-1757	37
hasło (n) ‘password’ (loan)	1854-1909	34
wiosna (f) ‘spring’	1942-2009	32
kolejka (f) ‘line’	1942-2009	32
pięciolatka (f) ‘five-year term’	1942-2009	32

As we can see from the data summarized in (44), the majority of the alternating nouns have *k* as their stem-final consonant (60%), which in all these words happens to be a part of the *(e)k* suffix, either in its nominalizing function (e.g. *setka* ‘a hundred’, *pięciolatka* ‘five-year term’), or in the productive diminutive function (only *kółko* ‘circle-dim.’ and *miasteczko* ‘town-dim.’), or as a frozen diminutive suffix (e.g. *matka* ‘mother’, *książka* ‘book’, *łóżko* ‘bed’, *córka* ‘daughter’, *butelka* ‘bottle’<sup>46</sup>). Two other relatively frequent sequences are *CwV~Cew* and *CnV~Cen*, each with 16% occurrences; almost all of these nouns contain an etymological suffix, which is hardly transparent or not transparent at all. There are also two monosyllabic nouns in this group.

(44) The summary of clusters in *e~∅* feminine and neuter nouns ranking 1-2009.

type of cluster	ranks 1-1002	ranks 1003-2009	total ranks 1-2009
CkV~Cek	3	12	15 (60%)
CwV~Cew	2	2	4 (16%)
CnV~Cen	2	2	4 (16%)
#grV~g’er		1	1
#dnV~den		1	1
total nouns:	7	18	25

Nouns with the fixed *-CC* sequence are much more numerous in these frequency ranges, as shown in Tables XXVII and XXVIII<sup>47</sup>, with the results summarized in (45). Similarly to the previously discussed masculine nouns, but even to a greater extent, the types of clusters in *e~∅* alternating and *CC* non-alternating nouns are largely complementary, which can be clearly seen on a closer examination. Even though *k* as the final consonant is found among non-alternating nouns, in six cases (out of nine) it is a part of the neutral suffix *-(i)sko* which always has the fixed form – thus there are no comparable examples in (44). Only one instance of the non-alternating sequence *lk* (*walka* ‘fight’, gen. pl. *walk*) has a correspondent in the alternating class (*butelka* ‘bottle’, gen. pl. *butelek*). The final *n* occurs three times in (45) and is a part of the never-alternating feminine suffix *-yzna* (*mężczyzna* ‘man’, *powierzchnia* ‘surface’).

<sup>46</sup> Most of these nouns occur in suffixless, etymologically basic forms, reinterpreted as augmentatives, e.g. *księżka* ‘book-augm.’ *łóżka* ‘bed-augm.’, *córka* ‘daughter-augm.’, *butla* ‘bottle-augm.’ (etymological base *mac* ‘mother’ is retained only in some frozen swearwords).

<sup>47</sup> I do not include nouns in which gen. pl. ends in *-i*, cf. *sytuacja* ‘sytuacja’ gen. pl. *sytuacji* etc.

Table XXVII. Nouns of feminine and neuter genders in the first 1000-word list ending in -CC in gen. pl.

państwo (n) ‘country’	152	332
liczba (f) ‘number’	211	249
przedsiębiorstwo (n) ‘enterprise’	239	228
walka (f) ‘fight’	256	218
prawda (f) ‘truth’	300	189
warstwa (f) ‘layer’	341-343	169
forma (f) ‘form’ (loan)	349-351	164
województwo (n) ‘adm. region’	349-351	164
stanowisko (n) ‘position’	359-361	161
zjawisko (n) ‘phenomenon’	387-391	148
gospodarstwo (n) ‘household’	507-510	119
rolnictwo (n) ‘agriculture’ <sup>48</sup>	511-513	118
wojsko (n) ‘army’	522-527	116
mężczyzna (m) ‘man’	558-564	110
służba (f) ‘service’	602-613	103
nazwisko (n) ‘family name’	673-684	93
bezpieczeństwo (n) ‘security’	765-778	83
środowisko (n) ‘environment’	788-799	81
ministerstwo (n) ‘ministry’	800-810	80
towarzystwo (n) ‘company’	820-835	78
nazwa (f) ‘name’	866-877	74
zwycięstwo (n) ‘victory’	893-903	72
usta (n-pl. tantum) ‘mouth’	914-930	70

Table XXVIII. Nouns of feminine and neutral genders ranked 1003-2009 with -CC in gen. pl.

przerwa (f) ‘break’	1019-1036	64
reszta (f) ‘rest, change’ (loan)	1047-1965	62
kadra (f) ‘personnel’ (loan)	1148-1164	56
pismo (n) writing	1148-1164	56
rezerwa (f) ‘reserve’ (loan)	1224-1243	52
sekunda (f) ‘second’ (loan)	1224-1243	52
lampa (f) ‘lamp’ (loan)	1244-1270	51
wyspa (f) ‘island’	1244-1270	51
mistrzostwo (n) ‘mastery’	1322-1355	48
święto (n) ‘holiday’	1380-1413	46
mięso (n) ‘meat’	1414-1443	45
szansa (f) ‘chance’ (loan)	1414-1443	45
gwiazda (f) ‘star’	1444-1475	44
siostra (f) ‘sister’	1541-1570	41
troska (f) ‘worry’	1541-1570	41

<sup>48</sup> This word, as well as *bezpieczeństwo* ‘security’, is not normally used in plural.

klęska (f) ‘disaster’	1571-1612	40
lotnisko (n) ‘airport’	1571-1612	40
dobro (n) ‘good’	1613-1648	39
reforma (f) ‘reform’ (loan)	1613-1648	39
izba (f) ‘chamber’	1657-1697	38
płaszczyzna (f) ‘surface’	1657-1697	38
norma (f) ‘norm’ (loan)	1698-1757	37
ojczyzna (f) ‘homeland’	1698-1757	37
wydawnictwo (n) ‘publishing house’	1758-1805	36
krzywda (f) ‘injustice’	1806-1853	35
małżeństwo (n) ‘marriage’	1806-1853	35
taśma (f) ‘tape’ (loan)	1854-1909	34
prośba (f) ‘request’	1910-1941	33
przestępstwo (n) ‘crime’	1910-1941	33

(45) The summary of feminine and neuter nouns ranking 1-2009 with –CC in gen. pl.

type of cluster	ranks 1-1002	ranks 1003-2009	total ranks 1-2009
stf	8	3	11
sk	5	3	8
Ct (Cd)	2	3	5
tstf	3	1	4
Cm	0	4	4
Cb	2	2	4
Cv	1	2	3
zn	1	2	3
Cp	0	2	2
Cr	0	2	2
ns	0	2	2
nt (nd)	0	2	2
lk	1	0	1
str	0	1	1
total nouns:	23	29	52

For completeness, *CeC* stem-final feminine and neuter nouns should be considered. I have spotted 16 nouns among 1-2009 frequencies, some native, e.g. *rzeka* ‘river’, *potrzeba* ‘need’, *drzewo* ‘tree’, and some borrowed, e.g. *gazeta* ‘newspaper’, *zero* ‘zero’. Nouns of this type are rather sporadic among all frequency ranges and do not seem to be of any value as possible extendable patterns. Since they do not interfere with analogical processes affecting (or triggered by) *e~∅* alternating nouns, they will be excluded from further discussion in order not to complicate the issue, which will turn out to be rather complex by itself.

In lower frequencies, feminine/neuter alternating nouns are much more numerous and the percentage of *k*-final stems is even more prominent than for high ranks. This

result is partly due to the very productive derivation with *k*-final feminine suffixes, but it also correlates with other facts discussed below. In the group of four-occurrence words of ranks 8739-10355 in *Słownik* (1990), as many as 104 (92,9%) out of 112 alternating nouns of feminine (total 98) and neuter (total 14) genders have a stem-final *-k*, e.g. *omyłka* ‘error’, *ramka* ‘frame-dim.’, *pocztówka* ‘post-card’, *drzewko* ‘tree-dim.’, *lusterko* ‘mirror-dim.’. Only seven feminine nouns have other consonants in this position: two instances of *n* (e.g. *wanna* ‘bath’) and *v* (e.g. *łyżwa* ‘skate’), one *l* (*szabla* ‘saber’), one *w* (*perła* ‘pearl’) and one monosyllabic word *ćma* ‘moth’, and there is only one neuter noun (*wiadro* ‘bucket’). The distribution of *e*~ $\emptyset$  alternating feminine/neuter nouns looks particularly interesting when compared to the distribution of masculine nouns in the same frequency range, shown in (46) below. We can see that although feminine/neuter nouns are 2.5 times more numerous, they comprise exactly the same number of different types (seven, counting each monosyllabic noun as a single type), but are much more homogenous due to the extremely large proportion of *k*-final stems and few occurrences of other types (in fact fewer than in masculine nouns). This effect does not seem to be coincidental, but follows from two biases among feminine/neuter nouns: a positive bias towards *k*-final alternating stems<sup>49</sup> and a negative bias against many other *C*-final alternating stems, which I will discuss in the following section.

(46) The distribution of *e*~ $\emptyset$  alternating nouns ranking 8739-10355.

	masculine	feminine	neuter	f. & n. jointly
<i>k</i> -final stems:	30 (68,2%)	91 (92,9%)	13 (92,9%)	104 (92,9%)
other <i>C</i> -final stems:	14 (31,8%)	7 (7,1%)	1 (7,1%)	8 (7,1%)
	<i>r</i> (6)	<i>n</i> (2)	<i>r</i> (1)	
	<i>l</i> (3)	<i>l</i> (1)		
	<i>ts</i> (2)	<i>v</i> (2)		
	<i>t</i> (1)	<i>w</i> (1)		
	<i>mech</i>	<i>ćma</i>		
	<i>bez</i>			
<i>C</i> -final types:	7	6	2	7
total nouns:	44	98	14	112

#### 4.3.2. Multifaceted analogy in feminine and neuter nouns

Recall that in feminine/neuter nouns the minor *CeC* allomorph occurs in one, rather infrequent, word-form of the paradigm and is thus on the whole much less salient than in the case of masculine nouns with a more balanced distribution of the two allomorphs. Consequently, it will be more susceptible to either pattern or stem analogy. The former is clearly available in the case of *k*-final stems. Nouns of this type are numerous at various frequency ranges due to the existence of very productive suffixes. They are found among the most frequent nouns and, more importantly, they occur in the plural, including the

<sup>49</sup> With the exception of the aforementioned non-alternating neuter suffix *-isko*.

word-form of the genitive plural, recall Tables XXV and XXVI with common nouns as ‘mother’, ‘book’, ‘daughter’, ‘song’ etc. The  $e\sim\emptyset$  alternation of  $k$ -final stems can then be successfully memorized for a number of high-frequency nouns and, due to the existence of a large class of such nouns (of various frequency ranges) it becomes a strong morphophonemic pattern, especially for feminine nouns. Therefore, the original historical alternation is not only easily continued, but may spread onto other lexical items of a similar phonological, but not necessarily morphological, make-up, cf. loanwords such as *maska*: *masek* ‘mask-nom.: gen. pl.’, *marka*: *marek* ‘mark (currency); brand-nom.: gen. pl.’, etc.

However, the  $Ck\sim Cek$  alternation does not have an obligatory character for all consonants in the position of  $C$ . Recall that among ranks 1-2009, three  $Ck$  non-alternating feminine nouns have been found, one with the  $lk$  cluster (*walka* ‘fight’) and two with the  $sk$  cluster (*troska* ‘worry’ and *klęska* ‘disaster’). These are precisely the combinations which are distinguished by Laskowski (1975:43) in his summary of alternating versus non-alternating clusters as those which may but do not have to alternate, although such nouns are relatively rare (apart from the fairly frequent neuter nouns suffixed with  $-isko$  discussed later). In terms of the present analysis, the rarer and historically never alternating (due to the absence of yer)  $-lk$ ,  $-sk$  endings of the genitive plural, constitute a weak pattern. This explains why some loanwords keep the invariant  $Ck$  sequence throughout the paradigm or have two variants in variation. For example, *kalka* ‘carbon paper’ forms the genitive plural as *kalk* or *kalek*. The first option is more natural for a loanword (“be as much as you are, if you can”); the second one complies with the overwhelming pattern (“be as everybody else”).

The strong pattern of alternating  $-C(e)k$  clusters sharply contrasts with the recessive character of alternations involving clusters with other consonants in the stem-final position. The differences are especially striking if we consider the fact that some of these consonants are remnants of historically yer-initial suffixes, hence their resemblance to  $k$ -final stems is not only phonological, but also morphological. This is the case of the four classes of feminine nouns discussed below.

Among the feminine nouns of frequencies in the range of 1-2009 listed earlier in Tables XXV-XXVIII, some end (in the nominative) in  $-da$ ,  $-ba$ ,  $-na$  and  $-va$ , which in the majority of cases go back to unproductive historical suffixes  $-ьda$ ,  $-ьba$ ,  $-ьna$  and  $-(t)ьva$ . The few high frequency nouns which have appeared earlier in the data will be discussed in the larger context of the distribution of the particular types in the Polish lexicon, with Witold Doroszewski’s *a tergo* dictionary (*Indeks* 1965)<sup>50</sup> as a source.

Three nouns with the final  $-da$  occur in the lists of non-alternating high-frequency nouns. One of them, the underived *gwiazda* ‘star’, did not contain a historical yer, hence was never alternating; two others are etymologically derived: *prawda* ‘truth’, *hist.* ‘integrity’ (cf. *prawy* ‘right’) and *krzywda* ‘harm’, *hist.* ‘immorality’ (cf. *krzywy* ‘crooked’, *hist.* ‘immoral’). Notice that both nouns have abstract meanings in the

---

<sup>50</sup> It should be noted that the index is based on a rather old source, viz. S. B. Linde’s dictionary from the 1850s and some words are completely or almost out of use. (Michalewski (1984) contains a similar index based on newer sources, but it includes only suffixes synchronically motivated and is not exhaustive.) In my later count, I will exclude proper names (except for common Christian personal names), different phonological variants, as well as more than one occurrence of compounds (unless they are lexicalized with unpredictable meanings).

contemporary language and had such meanings in the past, too. At present, their usage in the plural form is rather limited (although not impossible); in this case the *-Cd* ending is used in the genitive plural. The nouns were not much used in the plural in their previous, past meanings. In the dictionary of Old Polish (*Słownik* 1968-2004), there are few occurrences of the genitive plural *krzywd* (in this, non-alternating form only), but no cases of the genitive plural of *prawda* (among as many as 4285 attested occurrences of the word).

*Indeks* (1965) contains almost a hundred nouns ending in *-Cda*, but with the exception of the two just mentioned (and the negated *nieprawda* ‘untruth’), practically only the arguable *jazda* ‘driving’ (*hist.* ‘travel’) and the obsolete *bajda* ‘nonsense’ (according to Bańkowski 2000, a 19<sup>th</sup> c. innovation) actually have the identifiable suffix *-da*. The extreme rarity of derived nouns results from the fact that the suffix goes back to the Proto-Slavic period and was not much productive at a later time. Apart from twelve nouns with the stem nasal vowel (e.g. *kołęda* ‘Christmas carol’) and very few other native ones (e.g. *pogarda* ‘disrespect’), all nouns ending in *-Cda* are loanwords, e.g. *banda* ‘gang’, *giełda* ‘stock market’, *komenda* ‘command’, *sekunda* ‘second’, etc. The nouns which may occur in the plural always take the non-alternating form *-Cd* in gen. pl.

Let us observe that the difference in behavior between *-C(e)ka* nouns, alternating in their great majority, and never alternating nouns ending in *-Cda*, relates strictly to frequency. In the past, both classes were potentially alternating, due to the presence of the stem-final historical yer. In the case of the latter class, however, the *-Ced* variants were almost never revealed, due to the rarity of occurrence of the genitive plural form of such nouns in actual language usage. The former class, on the contrary, was large in the past and is still productive. The *-Cek* variant had (and still has) a good chance of being memorized, since many nouns in this class, including very common ones, denote objects and persons and are not restricted as to their occurrence in the plural genitive.

An alternative analysis of the distinction between these two classes could appeal to phonology and the argumentation that *-Cd* codas are better tolerable than *-Ck* ones. Although there is in principle nothing wrong with such reasoning, it would be quite unappealing in the case at hand. First, *-Ck* codas freely occur in other morphophonological contexts, as already exemplified in this chapter. Secondly, evidence of three other historical suffixes, which will be discussed momentarily, complies very well with the frequency-based explanation. It will be demonstrated that nouns ending in *-ba*, *-na* and *-va* situate themselves between the two extremes of *-ka* and *-da* nouns in terms of frequency and in terms of their eagerness to alternate.

Out of the four previously listed words ending in *-ba*, with the exception of *izba* (< \**istba*) ‘chamber’ of rather uncertain etymology, three are derived: *liczba* ‘number’ (cf. *liczyć* ‘to calculate’), *służba* ‘service; servants’ (cf. *służyć* ‘to serve’) and *prośba* ‘request’ (cf. *prosić* ‘to request’). In the contemporary language, all these nouns take the *-Cb* ending in the genitive plural. However, in Old Polish they had abstract meanings, such as, respectively: ‘calculating, the process of counting’, ‘serving, being a servant’ and ‘(act of) requesting’. As such, they were not used in the plural. Bańkowski specifically says that *liczba* was used in Old Polish only in the singular (2000 vol. 2:39), and that *prośba* had plural forms enforced by the translation of the Bible and not occurring in the speech before the 19<sup>th</sup> century (2000 vol. 2:789). It can be safely assumed that various other nouns suffixed with *-ba* which had regular, transparent semantic derivation as *nomina*

*actionis* also lacked plural forms. This appears to be true for nouns of rather limited current usage, such as *chwalba* ‘(act of) praising’, *kośba* ‘(act of) scything’, *młóćba* ‘(act of) threshing’, *siejba* ‘(act of) sowing’. Some nouns underwent semantic changes to designate more concrete objects and as such, they are free to pluralize, e.g. *strzelba* ‘rifle’, *hist.* ‘(act of) shooting’, *rzeźba* ‘sculpture’, *hist.* ‘(act of) slaughtering’. Few nouns are early loanwords, e.g. *farba* ‘paint’, *korba* ‘crank (handle)’ and *torba* ‘bag’.

The *-Cba* ending nouns constitute a small category in the contemporary language. Those mentioned above with the addition of several other nouns of lower frequencies practically exhaust the list. *Indeks* (1965) contains altogether about fifty nouns of this shape (including four in *-ęba*), but many of them are obsolete. All nouns which may take the plural have the invariant *-Cb* ending, except for the common noun *torba* ‘bag’, whose genitive plural is based on the alternating pattern *toreb*. It is probably not coincidental that the alternation was preserved in just this word. Or, to put it differently, if the alternation were to be preserved somewhere, it would be most likely in this particular word. Firstly, this noun always had in the past, and still has, a concrete meaning, so that the genitive plural case had a real chance to occur. Secondly, the word has always belonged to common, everyday vocabulary, so that the alternation could be easily memorized.

The historical suffix *-bna* is synchronically hard to recognize as a suffix at all. Among the high frequency data, three feminine nouns have been spotted: *wojna* ‘war’ (cf. *wojsko* ‘army’), *wiosna* ‘spring’ (synchronically non-compositional, cf. Lithuanian *vasara* ‘summer’, *pavasaris* ‘spring’, Latin *vēr* ‘spring’, etc.) and *panna* ‘maiden, Miss’ (cf. *pan* ‘master, Mr.’). All these nouns take *-Cen* variant in gen. pl., cf. *wojen*, *wiosen*, *panien*. Among lower frequencies of 8739-10355 rank in *Słownik* (1990), three nouns with the stem-final *-Cn* are found: *wanna* ‘bath’ and *rynna* ‘drainpipe’, which take the alternating gen. pl., cf. *wanien*, *rynien*, respectively, and *fontanna* ‘fountain’ with the non-alternating gen. pl. as *fontann*.<sup>51</sup> These data are fairly representative of the general facts which can be summarized as follows. The *-Cn* stem-final feminine nouns constitute a relatively small category with a moderately recessive character of the alternating pattern. About eighty of them appear in *Indeks* (1965), but 25 in this group are rare diminutives suffixed with *-(u)chna* (e.g. *babuchna* ‘grandma’, *wiochna* ‘village-dim.’, but mostly proper names, such as e.g. *Kachna*, *Kasiuchna*, *Kasiniuchna* – all of these included!) and 30 are suffixed with the obsolete suffix *-ówna* (e.g. *aptekarzówna* ‘pharmacist’s daughter’, *carówna* ‘tsar’s daughter’, etc.). Excluding these two groups and obsolete nouns, about twenty *-Cn* stem-final nouns are in actual language use, with a little more than half of them having plural forms. An almost exhaustive, representative list of alternating nouns (but with the exclusion of those which do not occur in the plural) is shown in (47) with the PWN Corpus figures of their WORD and genitive plural forms. Examples of non-alternating nouns are included in (48).

<sup>51</sup> I have excluded one more noun which appeared in this group, because of its exceptional adjectival declension, cf. *księżna* ‘duchess’, gen. pl. *księżnych*. For the moment, I also exclude nouns suffixed with *-izna* (*-yzna*) to which I will come later in this section.

(47) Nouns with the genitive plural in *-Cen*

WOJNA 16883, wojen 586 'war'  
WIOSNA 2924, wiosen 22 'spring'  
TRUMNA 766, trumien 39 'coffin'  
PANNA 2219, panien 113  
WANNA 425, wanien 19 'bath'  
RYNNA 188, rynien 33 'drainpipe'  
SOSNA 446, sosen 69 'pinetree'  
WEŁNA 372, welen 10 'wool'  
SARNA 282, saren 19, sarn 39 'roe deer'

(48) Nouns with the genitive plural in *-Cn*

KOLUMNA 1145, kolumn 127 'column'  
FONTANNA 333 fontann 0, 'fountain'  
SAWANNA 47, sawann 8 'savannah'

Laskowski (1975:40) claims that there is a clear opposition between native and foreign lexicon with respect to the alternation: while it is ubiquitous with native words (or at least optional with some, e.g. *sarna* 'roe deer', gen. pl. *saren* or *sarn*), it is absent in loanwords. While this observation is basically correct, there seems to be a tendency to eliminate the alternation in very rare native words, too. For example, in my own speech, I would probably do it in the case of diminutives suffixed with *-(u)chna*, hence *córuchn* 'daughters' as the genitive plural form instead of the dictionary *córuchen*, etc. Children also tend to level the stem in less frequent words. For example, an eleven year old girl interviewed by me used the correct alternating genitive plural forms of words, such as *wanna* 'bath-tub', *panna* 'maiden' or *trumna* 'coffin', as well as the correct non-alternating plurals of loanwords, such as *sawanna* 'savannah', *fontanna* 'fountain' or *kolumna* 'column', but regularized the stem in the less common *rynna* 'drainpipe' (i.e. *rynn* instead of standard *rynien*). A similar, weak tendency for leveling of the genitive plural is observed in some neuter nouns ending with *-Cno* (they are rarer than feminine nouns, only one, *okno* 'window', appeared in the high frequency lists). If, for example, I were forced to use rare words, such as e.g. *łajno* 'dung' or *gumno* 'barn (obs.)' in the genitive plural, I would rather employ the non-alternating stem variant instead of their dictionary forms *łajen* and *gumien*, respectively.

A comparison of the last two classes results in a straightforward conclusion. Even though the class historically suffixed with *-bna* was presumably never larger in size than the class historically suffixed with *-bba*, it preserved the alternation better, since many nouns belonging to it could occur in the plural and the minor case of the genitive plural was more salient.

The fourth feminine ending, *-Cva*, goes back to two etymological suffixes, *-tɔva* and *-ɔva*. The former regularly derived abstract nouns, e.g. *gonitwa* 'chase, race' (cf. *gonić* 'to chase'), *modlitwa* 'prayer' (cf. *modlić się* 'to pray'), *klątwa* 'anathema' (cf. *wykłąć* 'to curse'), *bitwa* 'battle' (cf. *bić się* 'walczyć'), etc. The latter derived concrete

nouns, rare in the contemporary language, e.g. *poszwa* ‘quilt cover’ or *podeszwa* ‘(shoe) sole’, both derived from verb *szyć* ‘to sew’ with different prefixes. There are also few non-derived loanwords with this ending, e.g. *tratwa* ‘raft’, *barwa* ‘color’, or *rezerwa* ‘reserve’. The last noun, as well as two others, *przerwa* ‘break’ and *warstwa* ‘layer’, appeared among our *Słownik* (1990) data in the non-alternating category. Altogether, the class of nouns ending in *-(t)wa* is slightly larger than each of the two previously discussed ones. *Indeks* (1965) contains over sixty nouns, the majority of which is still in use, although sometimes with very limited frequency. Most words of this category have the invariant *-Cv* genitive plural, as in all the examples in (49) below listed together with their PWN Corpus frequencies of WORD (in the increasing order) and genitive plural forms (included only if attested at least once in the corpus). It can be observed that most of these nouns are of rare occurrence and their genitive plural forms have particularly low frequency. Merely the last four are high-frequency nouns. Among them, *barwa* ‘color’ (obs. ‘paint’) was borrowed in 15<sup>th</sup> c., *rezerwa* ‘reserve’ is a new loanword, and *nazwa* ‘name’ is a 19<sup>th</sup> c. innovation (replacing earlier *nazwanie*, cf. Bańkowski 2000, v.1:287). Only *modlitwa* ‘prayer’ seems to be exceptional in eliminating the alternation in spite of being a high frequency noun. The dictionary of the 16<sup>th</sup> c. Polish (*Słownik* 1968-2004) already notes more uses of the new genitive plural *modlitw* (90 occurrences) than the old one *modlitew* (20 occurrences), out of the total of 2544 occurrences of this word<sup>52</sup>.

(49) Nouns with the genitive plural in *-Cw*

TYKWA 5, tykw 1, ‘gourd’  
 PARDWA 10, ‘grouse’  
 MAŁTWA 16, ‘cuttlefish’  
 KUTWA 17, ‘cheapskate’  
 PIGWA 18, ‘quince’  
 SAKWA 48 (obs.), ‘bag’  
 KUROPATWA 56, kuropatw 12 ‘partridge’  
 BULWA 87, bulw 8 ‘tuber’  
 ŻUCHWA 132, żuchw 1, ‘mandible’  
 GONITWA 138, gonitw 18 ‘chase, race’  
 KŁĄTWA 157, kłąt 12, ‘anathema’  
 LARWA 163, larw 49 ‘larva’  
 PŁETWA 250, płetw 60 ‘fin’  
 MODLITWA 1027, modlitw 79, ‘prayer’  
 BARWA 1807, barw 369, ‘color’  
 REZERWA 2943, rezerw 1037, ‘reserve’  
 NAZWA 7324, nazw 370, ‘name’

The alternating form of the genitive plural is limited to several nouns listed in (50). According to dictionaries, few of these nouns may have a non-alternating variant as well. But in actual language use, the variation is greater. With the exception of *\*łyżw*, which I have never heard, as opposed to *łyżew*, virtually all the rest of these nouns are

<sup>52</sup> Perhaps the unexpected leveling in this particular noun relates to its unusual (for a high frequency word) length of three syllables, which gets reduced to two in *modlitw*, cf. chapter 6.

acceptable in the leveled form of the genitive plural, at least to some speakers. Younger persons definitely tend to use the invariant form of the stem more than older ones and more readily with less frequent nouns (e.g. *poszw*, *tratw*, *pluskw*). But in the case of mid-frequency nouns, different speakers make different judgments. For example, among several persons of about thirty years of age, whom I interviewed, some preferred *bitew* but *brzytw*, while others just the opposite: *bitw* but *brzytew*. There is also a lot of hesitation among speakers as to which form should be the “correct” one, as well as acceptance of both variants of some nouns. Variation and the tendency for leveling is quite understandable in the case of this class, since alternating nouns do not occur at all among the highest frequencies. Thus, the *-Cew* pattern of the genitive plural is not supported by well memorized lexical items. This seems to be a reason why nouns of the *-(t)wa* class were in the past and are at present much more prone to leveling than the smaller *-na* class, which has at least some high frequency members, cf. earlier (47).

(50) Nouns with the genitive plural in *-Cew* (with a tendency for leveling)

POSZWA 8 ‘quilt cover’  
 TRATWA 107, *tratew* 2, ‘raft’  
 PLUSKWA 117, *pluskiew* 14 ‘bed bug’  
 BRZYTWA 124, ‘razor’  
 ŁYŻWA 125, *łyżew* 13 ‘skate’  
 LISTWA 129, *listew* 28, ‘slat’  
 PODESZWA 179, *podeszew* 2 ‘(shoe) sole’  
 KURWA 632 ‘whore’  
 BITWA 1237, *bitew* 77, ‘battle’

The above detailed discussion of five feminine noun endings has revealed that the nuances of behavior with respect to stem/pattern analogy highly correlate with various frequency criteria, such as the size of a class, non-vacuous representation of a class among the highest frequencies and, above all, non-vacuous representation of a minor alternant at high-medium frequencies. I have also argued that the phonological criterion of syllable well-formedness does not matter in the selection of the genitive plural *-CeC* or *-CC* ending. This claim, in addition to the arguments mentioned earlier, is further supported by the existence of non-alternating suffixes which may contain the same final consonant as some of the alternating (partly alternating) classes. I will return to this issue later in this section. Before that, I will briefly describe the remaining alternating classes of neuter and feminine genders.

In neuter nouns, the second most frequent (after *-Cko*) alternating ending is *-Cwo* (ort. *-Cto*). Four alternating nouns of this class were found among high frequency data (*światło* ‘light’, *źródło* ‘source’, *skrzydło* ‘wing’ and *hasło* ‘password’). Although the class is not too large (over a hundred members in *Indeks* (1965), but many of them obsolete or dialectal), the nouns freely occur in the plural and the alternation is supported by the existence of a moderately productive suffix *-(a)dło* which can derive new lexical items. The only exception seems to be *rzemiosło* ‘craft’ with the non-alternating genitive plural as *rzemiosł*. It should be noted that final *-Cł* clusters are tolerated well in Polish in other morphophonological contexts, e.g. in the nominative of several masculine nouns,

cf. *przemysł* ‘industry’, or in verbal inflection, cf. *part* ‘he pushed’, *wiódl* ‘he led’, *plótl* ‘he plaited’, etc.

Apart from the classes of various sizes distinguished above, alternations involving other consonants in the stem-final position are sporadic or limited to very small, unproductive classes. In some cases, the *-CeC* form of the genitive plural may look as if it were obligatory, since all words with a given final consonant exhibit it. But this impression is not quite accurate, as I will illustrate with the following examples.

Among the lower-frequency data of rank 8739-10355, one alternating neuter noun with the final *r* appeared, viz. *wiadro* ‘bucket’. This class covers several words only, e.g. *srebro* ‘silver’, *lustro* ‘mirror’, and we can apparently agree with Laskowski (1975:38), who claims that the alternation is obligatory for neuter nouns. But the situation is a little more complex. Very infrequent new loanwords with the stem-final *-r* do not readily accept the *-CeC* genitive plural. For example, *bistro* ‘bistro’ has the dictionary genitive plural as *bistrów* with the *-ów* ending adopted from other declension types. Speakers, tested by me as to the phrase *pięć X* ‘five Xs’, requiring the genitive plural of X, would rather say *bistr* (and never *\*bister*), commenting, however, that ‘*pięć knajp typu bistro*’ ‘five restaurants of the *bistro* kind’ sounds much more as what they would actually say. Let us observe that avoidance strategies, either by substituting a different genitive plural suffix or by eliminating this inflectional case all together, are well understood. The *-Cer* variant is not salient enough to create a productive pattern, applicable to new lexical items, but the *-Cr* variant does not constitute a pattern at all, since no lexical items exhibit it. Similar lack of confidence is found in the case of other rare (seemingly obligatory) alternating sequences, such as e.g. the feminine nouns ending in *-Cla* or *-Cwa*, which appeared among the data of rank 8739-10355 (each of them once). Although in the past, there existed limited cases of pattern analogy with such sequences (cf. *jodła* ‘fir’ discussed in the following section), there does not seem to be any productivity of it now. Speakers, forced to use the genitive plural of a new word as e.g. *fatwa* ‘fatwa’ occasionally heard on the news, uncomfortably produce *fatw*, judging *fatew* as completely impossible (cf. *miotel* ‘brooms-gen.’). Such examples and a few more of a similar kind involving other stem-final sequences clearly demonstrate that very small classes of alternating nouns occurring only among rarer frequencies have a recessive character.

Let us now return to productive but not alternating suffixes. I will briefly discuss those which have appeared in the data of the high frequency nouns (cf. earlier summary in (45)).

The most numerous non-alternating sequence results from the presence of the productive neuter gender suffix *-stfo* (ort. *-stwo*), with ten occurrences, which derives abstract nouns, e.g. *mistrzostwo* ‘mastery’, *zwycięstwo* ‘victory’, *bezpieczeństwo* ‘security’. The second most frequent suffix, *-isko*, with six occurrences, most often has an abstract or locative meaning, cf. *zjawisko* ‘phenomenon’, *środowisko* ‘environment’, *nazwisko* ‘family name’, *lotnisko* ‘airport’.<sup>53</sup> Likewise *-tstfo* (ort. *-ctwo*), with four occurrences, e.g. *rolnictwo* ‘agriculture’, *wydawnictwo* ‘publishing house’. The feminine suffix *-izna* (*-yzna*) occurs in three nouns in the data, namely, *mężczyzna* ‘man’, *ojczyzna* ‘homeland’ and *powierzchnia* ‘surface’. Even though these nouns may pluralize, they are

---

<sup>53</sup> An unrelated meaning is that of (rare) augmentatives, e.g. *psisko* ‘dog-augm.’.

rather exceptional in their category. Especially striking is the case of the first of these nouns, which is perfectly countable and has a very concrete meaning, but this is due to the unusual semantic changes the word underwent in the course of its history. It originally meant ‘being male’, which is semantically transparent, regularly derived by the abstract suffix from the adjective *męski* ‘male’. Later on, the word started to be used in the collective sense ‘men’ from which the contemporary singular meaning of ‘man’ was extended at a still later time. Apart from this exception and a handful of other nouns which may, but rather rarely do take a plural form (as the two other examples from the data or e.g. *trucizna* ‘poison’), nouns with the *-i(y)zna* suffix occur only in the singular, which correlates with their typically abstract (and occasionally collective) meaning. Representative examples include: *bielizna* ‘lingerie’, *siwizna* ‘gray hair’, *tężyzna* ‘fitness’, *francuszczyzna* ‘French style or language’, *niemcyszczyna* ‘German style or language’, *włoszczyzna* ‘Italian style or language’ with a further extension into ‘assorted combination of soup vegetables’.<sup>54</sup>

All of the above suffixes are very productive and comprise large classes (e.g. there are about 140 nouns with the *-i(y)zna* suffix included in *Indeks* (1965)). In the contemporary language, some of such nouns are used in the plural form which is mostly due to semantic changes which affected the original *abstracta*. It can be hypothesized that in the past, the usage of the plural was still more limited than at present. If ever these nouns had the alternating form of the genitive plural, it had no chance to survive. And since there is no pattern of alternation, new lexical items, even frequent, will not exhibit it, either.

#### 4.3.3. The problem of the Base in feminine and neuter diminutives

As in the case of masculine nouns treated earlier in section 4.2.3, inner stem allomorphy is impossible in feminine and neuter diminutives and the only alternation is that of the suffix *-(e)k* itself. Likewise, the generalized variant of the inner stem often consists of the [...CeC] allomorph. This of course would present a serious problem for a frequency-based analysis, since the leveled form coincides with the stem of the genitive plural – the only and relatively rarely used word-form of the base noun’s paradigm. As the examples in (51) illustrate (all of them familiar from the previous discussion), the base of the diminutive may itself be derived or not, native or borrowed. What seems to be even worse, the same [...CeC] inner stem shape appears in a significant number of diminutives whose base nouns have non-alternating *-CC* final stems, as exemplified in (52). Such cases are very problematic for any approach in which the *e-∅* alternation results from an underlying presence of a “ghost” vowel: it seems not to be there underlyingly, as it does not manifest itself in the base noun paradigm but it seems to be there, when diminutives are considered.<sup>55</sup> Both problems, of frequency, as well as of

<sup>54</sup> According to the tradition, it was Bona Sforza, an Italian wife of king Zygmunt I, who brought and popularized many new vegetables in 16<sup>th</sup> century Poland. Although a concept of *włoszczyzna* (usually sold in small packages) as a necessary base of almost every kind of soup is not unique to Polish people, I wonder how many other languages have a specific word for it. NB. *włoszczyzna* normally can not refer to vegetables used for other purpose than cooking a soup and other vegetables (than a prescribed basic set) can not be called *włoszczyzna*, even if cooked in the soup.

<sup>55</sup> In a recent OT-based paper, Jarosz (2005) recognizes this problem and postulates underlying yers in words as those of (52), together with a rather complex strategy of deleting them in a base noun paradigm,

underlying representation, cease to exist under a templatic account, to which I will return after pointing out still another problem.

(51) Diminutives of *e~∅* alternating feminine and neuter nouns

base noun: nom./gen.pl.		diminutive: nom.	gloss (base noun)
maska	masek	maseczka	‘mask’
książka	książek	książeczka	‘book’
torba	toreb	torebka	‘bag’
szabla	szabel	szabelka	‘saber’
wanna	wanien	wanienka	‘bath-tub’
perła	perł	perłka	‘pearl’
okno	okien	okienko	‘window’
wiadro	wiader	wiaderko	‘bucket’
lustro	luster	lusterko	‘mirror’
źródło	źródeł	źródełko	‘spring; source’

(52) *CeC*-diminutives of non-alternating *CC*-stems

base noun: nom./gen.pl.		diminutive: nom.	gloss (base noun)
taśma	taśm	tasiemka	‘tape’
forma	form	foremka	‘form’
palma	palm	palemka	‘palm-tree’
kolumna	kolumn	kolumienka	‘column’
cyfra	cyfr	cyferka	‘digit’
butla	butli	butelka	‘bottle’
cukiernia	cukierń	cukierenka	‘patisserie’
warstwa	warstw	warstewka	‘layer’
państwo	państw	państwewko	‘country’
wyspa	wysp	wyseпка	‘island’

The final set of data in (53) consists of nouns which keep the non-alternating *-CC* sequence throughout both paradigms, of the base noun and of the diminutive. It is important to acknowledge the existence of this group in order to understand that the previously suggested templatic inner stem [...*CeC*] requirement is not imposed on all diminutives with stems ending in *-CC*. Nouns of (53), as well as of (52), may belong to a native or borrowed lexicon and their morphological structure does not seem to connect them to a particular class, either. Let us also observe that the membership in a given class is not directly determined by a historical fact of having/not having a yer in the stem, since both kinds of nouns occur in all three groups.

---

but not in diminutives. But the author does not address a question, how these yer-containing nouns are different from yer-containing nouns of (51), which are not deleted in the genitive plural.

## (53) CC-diminutives of non-alternating CC-stems

base noun: nom./gen.pl.		diminutive: nom.	gloss (base noun)
karta	kart	kartka	‘card’
krosta	krost	krostka	‘pimple’
gwiazda	gwiazd	gwiazdka	‘star’
fałda	fałd	fałdka	‘fold’
bulwa	bulw	bulwka	‘tuber’
korba	korb	korbka	‘crank (handle)’
strzelba	strzelb	strzelbka	‘rifle’
farba	farb	farbka	‘paint’
małpa	małp	małpka	‘monkey’
różga	różg	różdzka	‘twig’ (dim. ‘wand’)
ciasto	ciast	ciastko	‘cake’

Out of the above three classes of nouns, the last one appears as the most straightforward, since no alternation is involved. On the other hand, the previous class of (52) looks as the most problematic of all. However, the two classes can be in principle reduced to one, if we observe that the distinction between them is largely based on phonological criteria. Specifically, the nouns in (52) tend to have a sonorant as a stem-final consonant (as in the first seven examples) or a cluster of at least three consonants (as in the following two examples). Very exceptionally (as in the last example), they have any other stem structure. On the other hand, nouns as in (53) have an obstruent in the stem-final position and, with extremely rare exceptions<sup>56</sup> never a sonorant. Consequently, nouns, as those in (52) can be treated as a sub-class of basically “non-problematic”, non-alternating nouns, as those in (53), with the only difference that a difficult medial cluster which would appear in diminutives of (52) is resolved by means of *e*-insertion. This process should not, in my opinion, be thought of as a simple phonological epenthesis, but rather as adoption of the [...CeC] template from another class of diminutives, such as (51). What remains to be said is why the nouns of (51) have a [...CeC] diminutive template and why this template has not been generalized to all diminutives (why not \**korebka*, given *torebka*?). Likewise, why are the diminutives of (51) not faithful to their more frequent [...CC] allomorph of the base noun, whenever syllable structure permits it (why not \**torbka*, given *korbka*?).

I will argue that the present situation with two competing diminutive templates connects to the parallel development of two kinds of diminutive stems, regularly derived from bases containing/not containing yers, which has led to a rather complex new reinterpretation, intertwined with syllable structure conditions.

There is sufficient historical evidence that the [...CeC] ending of the inner stem of diminutives directly related to the presence of the yer in the base noun, which regularly vocalized before the deleted yer of the diminutive suffix, cf. (54). With similar regularity,

<sup>56</sup> They include: the feminine *piosnka* as a rarer, marked variant of *piosenka* ‘song-dim.’ and the neuter *ziarnko* as a free variant of *ziarenko* ‘seed-dim.’ We can also add *ja[bwk]o* ‘apple’ (often simplified in colloquial language to *ja[p]ko*), for which the base non-diminutive form does not exist and very few masculine nouns with middle sonorants in cases other than the nominative, e.g. *czosnku* ‘garlic-gen.’

when the morphological base ended in a cluster of two consonants, the [...CC] inner stem structure appeared in the diminutive, since there was no base yer to vocalize before the deleted yer of the (*e*)*k* suffix, cf. (55). The quality of the stem-final consonant was at that stage irrelevant to the choice of the diminutive template, as examples in (54) and (55) illustrate, all with a stem-final sonorant of the inner base.

(54) Old Polish diminutives of *karczma* ‘inn’ and *okno* ‘window’

	Early Polish reconstruction	Old Polish	gloss (sg. nom.)
a/	kɾʲɛm-ɔka	kartʃem-ka	‘inn-dim.’
b/	ɔkɲn-ɔkɔ	ɔkʲen-kɔ	‘window-dim.’

(55) Old Polish diminutives of *bdła* ‘kind of mushroom’ and *pieśń* ‘song’

	Early Polish reconstruction	Old Polish	gloss (sg. nom.)
a/	bɔdw-ɔka	bedw-ka	‘mushroom-dim.’ <sup>57</sup>
b/	pesn-ɔka	pʲɔsn-ka	‘song-dim.’

Cases as those in (54) typically continue the [...CeC] template in diminutives, especially if the stem alternation is maintained in the base noun. Hence, for example, *torba* ‘bag’, with the genitive plural *toreb* and diminutive *torebka*, contrasts with *strzelba* ‘rifle’, with the genitive plural *strzelb* and diminutive *strzelbka* (although the latter stem historically contained a yer, too, cf. section 4.3.2). However, many nouns which have eliminated, or are in the process of eliminating the alternation in the base noun paradigm to the advantage of the [...CC] variant, still preserve the earlier [...CeC] stem variant in the diminutive. For example, many previously discussed nouns whose stems end in -*v* and which may have the -*CC* form in the genitive plural reject such a stem in the diminutive, cf. *brzytwa*, *brzytw~brzytew* ‘razor’, but *brzytewka* only (\**brzytwka*), *listwa*, *listw~listew* ‘slat’, but *listewka* (\**listwka*), etc. According to Laskowski (1975:30), the implication: “if there is an alternation (in his terms: epenthesis) in the base noun, then there is also epenthesis in the derivative, but not vice versa” holds almost unexceptionally. This result can be explained by appealing to the idea that the effort of producing a three-consonantal cluster word-medially is greater than the effort of producing a two-consonantal cluster word-finally. In addition, the pressure for leveling the diminutive stem in accordance with the stem of the base noun is, due to the semantic distance, smaller than the pressure for leveling the stem inside the base noun paradigm. If, for example, *toreb* and *torebka* are likely to become *torb* and *torbka* in the future (which is quite probable, since it is the only noun left as alternating in the -*Cba* class, cf.

<sup>57</sup> With the zero case ending of gen. pl., the predicted forms should have alternating inner stems, e.g. *kartʃmek* or *bdwek*, similarly to other examples of yer sequences. Unfortunately, because of low frequency of gen. pl. combined with low frequency of diminutives, it is practically impossible to find actual examples. In addition, it is likely that, because of their rarity, such words were the first ones to undergo stem leveling.

section 4.3.2), *torb* is expected to develop first with *torbka* following and not vice versa. Finally, the [...CeC] template in diminutives received further support from the independent innovation which affected diminutives of type (55) with the middle sonorant, to which I will turn now.

Many feminine and neutral stems inherited from Proto-Slavic ended in *-CN* sequences. In the diminutive form, after the loss of the suffixal yer, they surfaced with a difficult cluster with a middle sonorant, as previously illustrated in (55). In the contemporary language, only two nouns mentioned earlier survived as less common variants, cf. *piosenka~piosnka* ‘song’ and *ziarenko~ziarnko* ‘seed-dim.’ There is considerable historical and dialectal evidence that difficult sequences were often simplified by deletion of the offending sonorant, as exemplified in (56), cited after Ułaszyn (1956:44). In few cases, such forms survived until the contemporary standard language, cf. (56a). But in most cases they were replaced with new diminutives based on the [...CeC] template. Hence, in the contemporary standard language: *jodelka* ‘fir-dim.’, *sošenka* ‘pine-dim.’, *zwierciadelko* ‘mirror-dim.’, etc.

(56) Middle sonorant deletion (after Ułaszyn 1956)

a/ Remnant simplified diminutives in contemporary Polish

tarka	<	*tarłka	‘grater’
latarka	<	*latarnka	‘torch’

b/ Simplified forms attested in old texts (of limited use nowadays)

tarka	<	*tarnka	‘blackthorn’
ziarko	<	*ziarnko	‘seed-dim.’

c/ Simplified forms attested dialectally

betka	<	*bedłka	‘kind of mushroom’
jotka	<	*jodłka	‘fir-dim.’
soska	<	*sosnka	‘pine-dim.’
zwierciatko	<	*zwierciadłko	‘mirror-dim.’

Ułaszyn (1956) analyzes the process of *e*-insertion as a change of a primarily phonological character, which is triggered by a need to achieve greater salience, i.e. a full articulation of the sonorant trapped between two consonants and susceptible to loss. The emergence of the epenthetic vowel as a result of a gradient process, with a semi-syllabic sonorant as an intermediate stage is illustrated in (57), adapted from Ułaszyn (1956:59).

(57)

eg. *sosnka* > *sosn̩ka* > *sos<sup>ɛ</sup>nka* > *sošenka* ‘pine-dim.’  
 N > N̩ > <sup>ɛ</sup>N > eN

However, Ułaszyn recognizes some influence of forms with the “regular” *e* as a co-factor in this change, too.<sup>58</sup> My intuition is just the opposite and I would rather interpret the adoption of the [...CeC] template as a direct, non-phonological change, i.e. *sosnka* > *soseńka*, without any intermediate stages, since, if the process had a character of pure epenthesis, we would rather expect unattested \**sos[ɛ]nka* or \**sosn[ɛ]ka*, with the high vowel<sup>59</sup>.

In the particular case of diminutives, *e*-insertion appears as a better solution to the syllabification problem than the previously mentioned sonorant deletion, since it better maintains a correspondence relation between the inner stem of the diminutive and that of the base noun – an objective highly desirable in low frequency words. In the simplified *soska*, *jotka* etc., inner stems *sos-*, *jot-* lack a full segment when compared to the stems of their respective base nouns *sosn-* or *jodł-*. Such alternations are not found in Polish in other contexts, hence they do not constitute a memorizable pattern which could make the alternation more salient. On the other hand, the *e*-insertion does not only maintain the full segmental structure in the inner stem of the diminutive, but the two allomorphs are strongly connected by the existence of the common pattern of stem *e*~∅ alternations, being therefore easily identifiable as the same morpheme. It is also important to note that the adoption of a stem pattern already found among other diminutives reduces type allomorphy within a category of diminutives. Again, this is highly desirable in the case of a category of infrequent words (cf. Mańczak’s “differentiation law” from chapter 1). To conclude, forms, such as *soseńka*, *jodełka* etc. are in various respects better adjusted to the overall system of Polish than the respective simplified forms *soska* and *jotka*. It is also worthwhile to observe that the nouns which survived in contemporary Polish in the simplified form (cf. (56a) above) are those in which the loss of the sonorant is semantically irrelevant; either because of the very loose connection between the derived diminutive and its etymological base, cf. *latarka* ‘torch’ versus *latarnia* ‘(street) lamp post; lighthouse’, or, because the sonorant is not a part of the base word, cf. *tarka* ‘grater’ (from \**tarłka* with the unproductive derivational suffix *-ł*) versus *trzeć* ‘to grate’ (in other forms with the root allomorph *tar-*).

The historical process described above enforced the [...CeC] inner stem shape as a diminutive template giving it a power of a synchronic new constraint, which may come into conflict with faithfulness and output correspondence requirements, representing the “old” system. From the synchronic perspective, an interplay of all constraints is rather complex, as always in a situation of transition from one system to another. Below, I suggest a possible analysis, which complies with the basic OT architecture, as well as with the language use idea underlying the present work.

A rather obvious part of the argumentation involves a dominant constraint against CNC clusters (\*CNC), as well as a similar constraint against four-consonant clusters (\*CCCC), which will force epenthesis at the cost of necessary violations of

<sup>58</sup> He says: „[...] a given change is usually a result of a series of various factors, although of unequal decisive power, that is why I cannot say that formations with the „regular” *e* did not play a role of a co-factor in the phonetic development of the secondary *e* [...]” (Ułaszyn 1956:61, transl. from Polish I.K.S.).

<sup>59</sup> In this context, it is worthwhile to observe that Polish generally prefers morphophonological means to pure phonological epenthesis, which has cross-linguistic parallels as well. For example, Swahili uses a number of diversified strategies of “insert a dummy morpheme”-type to satisfy minimality requirement (cf. chapter 6) and almost no epenthesis, which is limited in this language to the adaptation of loanwords.

faithfulness/O-O correspondence (Faith/Cor), needed for cases, such as (52). In order to guarantee that the syllabification problem will be solved by means of vowel insertion and not (stem) consonant deletion, we could propose the ordered ranking of respective faithfulness constraints with Max dominating Dep. Generally speaking, such ranking is well motivated in the case of low frequency words (such as Polish diminutives), while the opposite ranking often characterizes the most frequent words (cf. chapter 6 for more discussion). However, in the particular case of non-phonological epenthesis observed here, a templatic constraint on the diminutive stem is needed anyhow. Therefore, I will assume that the attested form of the diminutive shape directly follows from it. The same constraint chooses the [...CeC] inner stem in the case of alternating nouns, such as those of (51). But there is a crucial difference between these two classes as to *how* the decision is actually made. In the latter case, the templatic constraint strictly dominates O-O correspondence which in the present approach is synonymous with saying that these particular stems belong to the  $e\sim\emptyset$  alternating category. In the former case of non-alternating nouns, it must be ranked below O-O correspondence whose violation is inevitable given the higher syllable structure constraints. When there is no conflict between syllable structure and O-O correspondence, as in the case of nouns in (53), the low templatic constraint plays no decisive role at all. In the present analysis, illustrated in (58), the problem of underlying “ghost” vowels ceases to exist, since the power of underlying representations as such is greatly reduced. It is quite irrelevant whether alternating nouns are assumed to have stems with an underlying *e* vowel subject to deletion in particular morphophonemic contexts, or whether this vowel emerges from nothing in other contexts, or whether the two allomorphs are underlying. Importantly, there is no need for highly abstract “yers”. The difference between alternating and non-alternating nouns (either with the *CC* stem-final sequence, as in e.g. *karta* ‘card’, or with the *CeC* sequence, as in e.g. *kareta* ‘carriage’) follows only from ranking of Cor above or below the templatic constraint.

There is one additional point that has to be made in reference to correspondence constraints. In the case of alternating nouns, there is always a violation of Cor with respect to one allomorph. In a noun, such as *torba* ‘bag’, gen. pl. *toreb*, the actual diminutive *torebka* contains an inner stem corresponding with *toreb* but not with *torba*, while a possible diminutive *\*torbka* would satisfy the Cor constraint in just the opposite way. If frequency criteria were not taken into account, these violations would be identically graded and the constraint would not make a difference in the evaluation at all. But in compliance with the present approach, usage correspondence violations, computed according to a sum of text occurrences of particular word-forms, lead to an emergence of the phonological Base. Since the stem allomorph *torb-* greatly outnumbers *toreb-*, the violation is more significant with respect to the latter and *torb* should be a Base for correspondence if it were determined by majority criteria. I indicate these unequal violations as \* and *n\**, respectively. Finally, how should the templatic constraint itself be formulated? Specifically, should each of the two allomorphs be made sensitive to phonological distinctions in their environments or to morphological distinctions only? The answer to this question is by no means obvious, since morphophonemic alternations are grounded in phonology to *some* degree only. I postpone a theoretical discussion of this problem until chapter 8. For the moment, I will assume a rough, morphology-sensitive formulation of the constraint (abbreviated as E- $\emptyset$ ), requiring the stem template

[...CeC] in the genitive plural of base nouns and throughout the diminutive stem. The mirror-image constraint on the [...CC] template in the remaining inflectional cases can be assumed, too (especially in a grammar with no URs), unless this major stem alternate is given as an underlying representation and is predicted by Faith.

The evaluation tableau in (58) demonstrates how the proposed constraint hierarchy predicts the correct optimal outputs for all kinds of nouns.

(58)

	*CNC, *CCCC	Cor {korba, butla, kareta etc.}	E~∅	Cor {torba etc.}
Base: korb- ☞ korbka korebka 'crank-dim.'		*!	*	
Base: butl- butlka ☞ butelka 'bottle-dim.'	*!	*	*	
Base: karet- ☞ karetka kartka 'ambulance'		*!	*	
Base: torb-, toreb- ☞ torebka torbka 'bag-dim.'			*!	n* *

In the above tableau, the E~∅ constraint refers to two morphophonological environments requiring the [...CeC] form of the stem, which reflects the historical fact of pure phonological realization of the yer in these contexts. However, in the contemporary reinterpretation of this constraint in templatic terms, the two environments do not seem to be equal, which is supported by evidence of a change in progress affecting some lexical items. Recall earlier exceptions (without a middle sonorant), such as *wyspa* 'island' with the genitive plural *wysp* and the diminutive *wyseпка*, as well as cases of variation in the base noun paradigm, but not in the diminutive (e.g. *listwa* 'slat', gen. pl. *listw~listew*, dim. *listewka*). Such cases suggest that the template is valued higher in the diminutive than in the base noun paradigm, or, that the correspondence within a paradigm is valued higher than the template. Consequently, two possible analyses can be proposed reflecting each of these two conceptualizations of the problem. Under the first account, the E~∅ constraint splits into two more specific ones, ranked differently with respect to all Cor constraints involving one stem, cf. the partial ranking in (59a). Under the second account, illustrated in (59b), particular correspondence constraints are ranked differently with respect to the templatic constraint, which in this case may have a more general formulation with regard to both environments. The former analysis stresses structural unity of the category of diminutives echoing Mańczak's differentiation law (cf. chapter

1), appropriately applying to low frequency lexical items. The latter analysis concentrates on subtle differences among correspondence constraints going hand in hand with various degrees of semantic strength among stem-sharing related words and highlights paradigm-internal closeness to a weaker cross-paradigm relation. This is also correct. Possibly, the most appropriate analysis should include both “splits”, as illustrated in (59c), assuming that some redundancy in grammar will not hurt.

(59) Partial ranking of “split” template and “split” correspondence constraints

- a/  $E\sim\emptyset$  {dim.} >> Cor { *wyseпка/wysp: wyspa* etc. } >>  $E\sim\emptyset$  {gen. pl.}
- b/ Cor { *wysp: wyspa* etc. } >>  $E\sim\emptyset$  >> Cor { *wyseпка: wyspa* etc. }
- c/ Cor { *wysp: wyspa* etc. } >>  $E\sim\emptyset$  {gen.pl.},  $E\sim\emptyset$  {dim.} >> Cor { *wyseпка: wyspa* etc. }

To close this section, it is perhaps worthwhile to point out a difference between the behavior of feminine and neuter diminutives vis-à-vis the less complex situation of masculine nouns, discussed previously in section 4.2.3. Recall, however, an “exception” found in that class, too, such as *wiatr*, *wiatru* ‘wind’ with no “ghost” vowel in the base noun paradigm, but with *e* in diminutive *wiaterek* (cf. section 4.1), which parallels the feminine/neuter class of (52). The extreme rarity of this pattern among masculine nouns (I cannot, actually, think of any other example except the colloquial *swetr* instead of *sweter* ‘sweater’, cf. section 4.2.1) presumably relates to a greater productivity of the  $e\sim\emptyset$  alternation in masculine nouns than in feminine/neuter ones, which sometimes led to reinterpretation of *-CN* final nouns as *CeN~CNV* alternating ones (cf. *ogień*, *ognia* ‘fire’, and not expected \**ogń*). In addition, some nouns with the non-alternating final *-CN* sequence derive diminutives not with *-ek*, but with the non-alternating suffix *-yk/-ik* (e.g. *teatr*, *teatru* ‘theatre-nom., gen., diminutive *teatrzyk*, *teatrzyku*), which is an option unavailable to feminine and neuter nouns.

## CHAPTER 5

# Semantic distance and contrast: differences between nominal and verbal paradigms

### 5.1. Why should verbs and nouns differ?

Generally speaking, analogy in verbal paradigms is triggered by the same frequency criteria as shown earlier for nominal paradigms. There exists abundant evidence in the literature proving that rarer verbal stems and patterns are replaced by more frequent ones. I briefly present a few examples below.

Mańczak 1996: 95n. (cf. also Mańczak 1958, 1978) points out analogical changes which affected verbal paradigms of French, Italian and Spanish. Even though each of these Romance languages underwent individual developments, not shared with others, there is a striking similarity in the general outline of changes. The more frequent a given tense (mode) is, the fewer analogical replacements, and *vice versa*, most analogical developments occurred in rare forms. For example, in the most frequent *indicativus praesentis*, one analogical form is found in the French paradigm, three in Italian and none in Spanish, but complete paradigms in all three languages have forms that developed analogously in the extremely rare *coniunctivus plusquamperfecti*.

In verbal paradigms, the form of the third person singular is the most “unmarked” in Greenberg’s (1966) terms and the most frequent. As Hock (1986, ch. 10) notes, this form is most resistant to analogical developments and may serve as a pivot for them. The author presents an example of such a change in Polish, where *jest*, the third person singular of the verb ‘to be’, became a basis of the whole paradigm (reinterpreted as a stem), with the exception of the unaffected form of the third person plural (Hock 1986: 221n.).

A frequent verbal category tolerates type allomorphy, while an infrequent one tends to be template-governed. Greenberg (1966:49) points out an example of Arabic, as well as some other Semitic languages, in which basic verbs have stems differentiated with respect to the first vowel – it may be *a*, *i* or *u*, but no such distinctions can be found in much rarer derived stems.

Similarly as in nouns, it may happen that a recessive alternation is maintained only in a few lexical items of the most frequent vocabulary, while a more transparent pattern characterizes a category as a whole. In Classical Arabic, stem vowel alternation was the only marker of the active/passive voice differentiation. In various modern varieties of Arabic, passive is usually marked more overtly by a prefix or an infix, while apophony is preserved in a few most frequent relic verbs and expressions, such as ‘it is said/ known/ found’ etc. (cf. Johnstone 1967, Retsö 1983). For example in Cairene Arabic, there are four such passive verbs and ten more other expressions borrowed from literary Arabic (Retsö 1983:91n). Another example of a similar kind can be found in Somali, in which an older type of prefixal verbal inflection (a complex pattern involving stem alternation) is limited to four frequent verbs (‘say’, ‘come’, ‘be-loc.’ and ‘know’) while all other verbs are inflected by more transparent suffixation (cf. Saeed 1987).

However, it is also well known that nouns and verbs exhibit asymmetries in various respects, including their behavior towards allomorphy/analogical leveling. In the remaining parts of this chapter, I will argue that such differences are often explainable in functional terms and follow from the fact that semantic distinctions coded by declensional cases are smaller than those coded by person and number (or tense, aspect, mood etc.) in verbal paradigms. Accordingly, since stem alternation maximizes an opposition, everything else being equal, it will correlate with verbal paradigmatic distinctions rather than with different nominal declensional cases (leaving aside specific situations when stem alternation is the only marker of a given category). Before discussing particulars resulting from this general reflection, I would like to justify a notion of semantic distance as relevant in this context.

The comparison of the semantic distance among nominal declensional cases and distinctions marked by verbal inflection may seem an impossible task, since the two are rather different kinds of objects. But we can relate them indirectly, pointing out certain diagnostics of their diverse intraparadigmatic distances.

Among languages of the world, morphological marking of verbal inflection is far more common than that of declensional cases. Since morphological marking reflects important oppositions, we may conclude that the oppositions expressed by the former are more significant than those expressed by the latter. The same reasoning explains the fact that suppletion is also more common in verbal paradigms than in nominal ones. To illustrate, several examples of fully suppletive (i.e. entirely distinct) stems are found among Polish verbs, either within a paradigm of one tense, or across tense/aspect, cf. *być* 'to be', *jestem* 'I am', *są* 'they are'; *idę* 'I go', *szłam* 'I went-imperf. (f)', *chodziłam* 'I went-imperf.-iterative (f)', etc. But only two nouns in Polish display complete suppletion between the singular stem and the plural stem (viz. *człowiek* 'human being' versus *ludzie* 'people' and *rok* 'year' and *lata* 'years') and it seems quite unthinkable that a suppletive form would be found within a partial paradigm of the same number.

Another argument comes from the observation that an opposition in declensional case does not necessarily accompany semantic differences, cf. Polish examples in (60) with near synonyms as (60a-b), or (60c-d), or (60e-g). Even though it can be claimed that each particular case has its own semantics (cf. Rudzka-Ostyn's 2000 cognitive study), it is by no means bluntly obvious, as illustrated by the examples in (60a) and (60e), (60c) and (60f), (60d) and (60h), which share the same case, as well as (60e) and (60h), which do not. As these examples show (and many others, e.g. active/passive structures), semantic distinctions carried by different cases are so small that it is easy to express the meaning conveyed by one case with a synonymous phrase in which a given noun will have some other case. Now, let us try to do the same with verbs and express the meaning of 'I do' with the verb meaning 'you do' or 'he does', or to express the concept of the past tense with the verb in the form of the future tense. This looks quite impossible<sup>60</sup>. And while it takes a linguist and a lot of thinking to find common semantics of a declensional case, every language speaker can understand and explain without any effort the sense of person, number or tense.

---

<sup>60</sup> Of course there are no impossible things in this world and we do find marginal examples, pragmatically strengthened, such as e.g. the Swahili expression *mwenzio*, which literary means 'your friend', but is often used in the sense of 'I'. Hence, *mwenzio anakupenda* 'your friend loves you' will actually be 'I love you'.

(60) Case similarities and differences

a/	patrzę na <u>drzewo</u> (acc.) 'I am looking <u>at the tree</u> '	e/	lubię (kocham) <u>fonologię</u> (acc.) 'I like (love) <u>phonology</u> '
b/	przyglądam się <u>drzewu</u> (dat.) 'I am looking <u>at the tree</u> '	f/	przepadam za <u>fonologią</u> (instr.) 'I love <u>phonology</u> '
c/	patrzę ponad <u>drzewem</u> (instr.) 'I am looking <u>above the tree</u> '	g/	podoba mi się <u>fonologia</u> (nom.) 'I like <u>phonology</u> '
d/	patrzę powyżej <u>drzewa</u> (gen.) 'I am looking <u>above the tree</u> '	h/	nie lubię <u>fonologii</u> (gen.) 'I do not like <u>phonology</u> '

The unequal semantic distance among members of nominal and verbal paradigms results in a number of consequences with regard to allomorphy and paradigmatic leveling, which are all triggered by verbs' relative tolerance or even preference for alternation and nouns' dissatisfaction with it. In a formal analysis, this effect can be attributed to the opposite ranking of stem correspondence constraints, as usual in cases of semantic differences. The following section will provide an illustration from Polish.

## 5.2. Noun-verb asymmetries with respect to analogy in Polish

All three alternations found in Polish nouns and discussed in chapters 2-4 have some correlates in verbal paradigms. I will limit the discussion to the *e~a/o* alternation, which has been shown to have a clearly recessive character in nominal declension. Remarkably, it seems to be fairly stable in verbs, even though it has a much narrower scope. The paradigms of the present and past tenses of the verb *nieść* 'to carry' in (61) illustrate the alternation which has remained in the original shape, according to its historical conditioning: *e* is found before a "soft" consonant and *o* before a "hard" one (underlined in (61)). The *e*-variant occurs in more forms of the present tense paradigm and the *o*-variant in more forms of the past tense paradigm. Within the present tense paradigm, the *o*-stem is found in the first person singular, which is cross-linguistically the second most frequent form, and in the third person plural. Altogether, this creates a fair representation of the minor pattern as to token frequency. In the paradigm of the past tense, none of the minor *e*-forms has high frequency, but the three of them create a sub-pattern of "plural masculine". Let us also note an *u*-variant found uniquely in the third person masculine singular of the past tense. The phonological conditioning of the alternation has also remained in other forms derived from the verb, cf. *niesi[e]nie* 'carrying', *niesi[ɔ]ny* 'carried', *zani[u]słszy* 'having carried', which I will not discuss.

(61) The present and past tense paradigms of *nieść* ‘to carry’

	present tense forms	past tense forms
1 sg	<u>niosę</u>	<u>niosłem</u> (m), <u>niosłam</u> (f)
2 sg	niesiesz	<u>niosłeś</u> (m), <u>niosłaś</u> (f)
3 sg	niesie	<u>ni[<u>u</u>]<u>sł</u></u> (m), <u>niosła</u> (f), <u>niosło</u> (n)
1 pl	niesiemy	<u>nieśliśmy</u> (m), <u>niosłyśmy</u> (f)
2 pl	niesiecie	<u>nieśliście</u> (m), <u>niosłyście</u> (f)
3 pl	<u>niosą</u> ‘I carry’ etc.	<u>nieśli</u> (m), <u>niosły</u> (f) ‘I carried’ etc.

The pattern illustrated in (61) is limited to nine stems only (Grzegorzczkova *et al.* 1984:79), but is further restricted in the past tense, to which I will come later. Most of the stems may occur with various prefixes modifying the lexical meaning, such as, *przy*, *prze*, *roz*, *wy*, *za*, etc. It is remarkable that one of the stems *wlok-* ‘drag’ is analogical, since the alternation occurs before a velar, which indicates that the pattern was not only stable, but even analogically extendable in a minimal fashion. Few of the verbs belong to the most frequent vocabulary of the first thousand in *Słownik* (1990): *brać* ‘to take’ with 162 occurrences, *jechać* ‘to go, drive’ with 91 occurrences and its derivative *przyjechać* ‘to come (by car)’ with 83 occurrences; few others are ranked within the second thousand of the most frequent words: *nieść* ‘to carry’, *przynieść* ‘to bring’ and *wynieść* ‘to take out’ (all containing the same stem *nos-*), and *pojechać* ‘go, drive (to)’. The remaining verbs have lower frequency.

The data in (62) contain PWN Corpus frequencies of particular word-forms and allomorph types of one of the most frequent verbs, *nieść* ‘to carry’, in the present tense. For the past tense, its perfective derivative *przynieść* ‘to bring’ has been used, since this form is more frequent than the imperfective. Still, we can see, that the occurrence of the minor *e*-allomorph in the past tense is very rare (in the absolute sense and relative to other allomorphs).

(62) PWN Corpus frequencies of (*przy*)*nosić* ‘to carry, bring’

Present tense occurrences:

NIOS – 237 (25,4%), NIES’ – 695 (74,6%)

(niosę 32, niosą 205, niesiesz 7, niesie 673, niesiemy 15, niesiecie 0)

Past tense occurrences:

PRZYNIOS – 1978 (60,5%), PRZYNIES’ – 146 (4,5%), PRZYNIOUS – 1145 (35,0%)

(przyniosłem 47, przyniosłam 32, przyniosłeś 9, przyniosłaś 5, przyniosła 865, przyniosło 379, przyniosłyśmy 0, przyniosłyście 0, przyniosły 641, przynieśliśmy 11, przynieśliście 2, przynieśli 133, przyniósł 1145)

The data in (63) show PWN frequencies of the third person plural masculine, the most frequent *e*-form of the past tense paradigm of the remaining five verbs of the category (three other verbs have partly suppletive, non-alternating stems: *jecha-* ‘go, drive’, *bra-* ‘take’ and *pra-* ‘wash’). We can see that with the exception of the first verb

on the list, *wieźć* ‘to transport’, all are extremely rare. To sum up, the alternating pattern in the past tense is limited to a category of six stems, out of which only two provide medium frequency forms of the *e*-allomorph; in the case of the remaining ones, this allomorph is not salient at all.

(63) PWN Corpus frequencies of 3 sg m. past tense forms of the remaining verbs

wieźli 38, przywieźli 155, zawieźli 57, podwieźli 7	‘transport’
wlekli 14, powlekli 14, zawlekli 9, przywlekli 0	‘drag’
gnietli 0, zagnietli 0	‘mash, knead’
pletli 0, zapletli 0	‘plait’
zamietli 0	‘sweep’

Taking into account the small size of this class of verbs and the difficult retrieval of the *e*-allomorph in the past tense, we could expect it to be leveled, especially in the light of the leveling observed earlier in nouns. Actually, considering the sizes of the two categories and their representation among the highest frequencies, we could expect the facts to be rather opposite to what they are. Recall from chapter 2 that nouns which are still alternating are better represented among high frequencies than the verbs here, not to mention the relatively large group of the nouns which are already leveled by analogy.

The surprising difference between the stable character of the *e~a/o* alternation in verbs in comparison to a recessive character of the alternation in nouns can be only explained by the fact that correspondence relations among various forms in the verbal paradigm are not as tight as in the case of the nominal paradigm. We can reflect this distance in the formal apparatus of OT as sketched in (64) below. As an illustration, I will use the stem *miot-* ‘sweep’, which is the only one shared by a verb of this class and a noun. *Miotła* ‘broom’ as a low frequency noun underwent leveling (as well as its derived adjective, cf. chapter 2), which is represented as the high ranking of its intraparadigmatic correspondence constraints. Similarly in the case of correspondence between the adjective and the noun. A greater distance among forms within the verbal paradigm, as well as a greater distance between less related stem-sharing words, such as e.g. the verb and the noun, comply with ranking of respective correspondence constraints below a templatic constraint.

(64) An OT analysis of the semantic distance in verbal versus nominal paradigms

Cor-N:N { *miotle: miotła* etc. }, Cor-Adj:N { *miotlasty: miotła* etc. } >> E~A/O >>  
 Cor-V:V { *zamietli: zamiotłam* etc. }, Cor-V:N { *zamietli: miotła* etc. }

The contrastive noun-verb behavior illustrated above for Polish cannot be of course observed in languages as e.g. English or modern Arabic which do not have nominal declension. However, even in languages like these, we can see a “verb effect”, i.e. a fact of maintaining an alternation in a verbal category in spite of its limited salience. A case of the so-called “strong” verbs provides a relevant example in English.

According to Bybee and Moder’s (1983) study (cf. also Bybee 2001, ch. 5), class *A* of the *sing: sang: sung* type is limited to eleven verbs only, while the list of class *B* of the *string: strung* type contains eighteen verbs (but four of them are non-standard

dialectal forms). In class *A*, a few verbs belong to the most frequent vocabulary (e.g. *come, begin, run*), but I can not see a single such item in class *B*, which consists only of mid and rare occurrence verbs (e.g. *win, spin, swing*). Still, as Bybee and Moder (1983) argue on the basis of historical and experimental evidence, the class is quite productive. In terms of the present analysis, this is possible because the tense/aspect distinction enhanced by the stem alternation constitutes a major semantic difference.

A similar example from Arabic, but involving inflection, will be discussed in the following section and was inspired by McCarthy's (2005) article.

### 5.3. Inflectional patterns in Moroccan and other Arabic dialects

The structure of a common type of the Moroccan Arabic noun is largely predictable by syllable well-formedness criteria and surfaces either as *CəCC* or *CCəC*. As McCarthy (2005) observes, no such conditions determine the template of the verb, which may only have the *CCəC* pattern (and never *\*CəCC*), regardless of the onset sonority. In McCarthy's analysis, this fixed template, which is found in the non-affixed third person masculine singular of the past tense, better complies with correspondence constraints within the whole verbal paradigm, computed according to a model of Optimal Paradigms. Looking at these data from a language use perspective, I will propose an alternative account of the Moroccan noun-verb asymmetry. I will argue that it is rooted in the historical development and correlates with such factors as category scope and semantic contrast and distance. The analysis can also explain some additional facts of modern Arabic dialects. I will start with a discussion of the historical source of the two nominal templates, taking Classical Arabic as a fairly good approximation of the ancestor language. The data of Moroccan Arabic (in a unified transcription) are cited after Caubet (1993) and Sobelman and Harrell (1963).

The loss of original Classical Arabic declensional suffixes reduced the shape of the noun to the mere stem in Middle Arabic, which led to the emergence of complex codas in nouns of the *\*CVCC-V(n)* type. Their reflexes in modern Moroccan often continue this pattern, with the only difference that the initial short vowel is reduced to [ə], as shown in (65).

(65) Moroccan *CəCC* nouns from Classical Arabic *CVCC* stems:

Moroccan		Middle Arabic	gloss
kəlb	<	kalb	'dog'
ʃəms	<	ʃams	'sun'
bənt	<	bint	'daughter'
zəld	<	zild	'skin'
nəfs	<	nafs	'soul'
fərq	<	farq	'difference'
qəwʃ	<	qawʃ	'arch'

The contemporary Moroccan type *CCəC* nouns originated in *CVVCV* stems of Classical Arabic, which reduced their structure as a result of syncope prohibiting short vowels in open syllables. The rule has a character of a synchronic constraint and is undominated in Moroccan (cf. McCarthy's (2005) constraint \*ə]σ), as well as in many other modern Arabic dialects. The role of syncope in Moroccan phonology can be compared to that which yer deletion played in Polish, since both rules operated without any regard for syllable structure and both brought about a series of unusual onsets, which often violate sonority hierarchy. We have seen such examples from Polish previously in chapter 4, the Moroccan data are shown in (66) below.

(66) Moroccan *CCəC* nouns from Classical Arabic *CVVCV* stems:

Moroccan		Middle Arabic	gloss
bgəɾ	<	baqar	'cattle'
ʕləm	<	ʕalam	'flag'
bʕəl	<	baʕal	'onions'
ʒməl	<	ʒamal	'camel'
dhəb	<	ðahab	'gold'
wtəd	<	watad	'peg'
ktəf	<	katif	'shoulder'

Recall also from the previous discussion of Polish that the final yer deletion (as well as some other processes including stem analogy) resulted in a number of sonority-violating final codas. Likewise in Moroccan: the unexceptional decay of declensional suffixes sometimes resulted in difficult word final sequences of two consonants in nouns of the *CVCC* type. Contrary to Polish, which tolerates such clusters, Moroccan chose to resolve the problem by adjusting such nouns to the *CCəC* template of the other type by means of word-medial metathesis. As illustrated by the examples in (67), the change affected mostly nouns with a final sonorant.<sup>61</sup> Again, we can see a clear parallel to some of the strategies chosen by Polish, such as for example the adoption of a diminutive template from another type of nouns discussed in section 4.3.3, which was also triggered by syllable structure criteria.

---

<sup>61</sup> But there are some cases of sonorant in this position, cf. *ħolm* 'dream', *ħokm* 'judgement' or *nəħw* 'grammar' (the last word could be perhaps a loan-word from literary Arabic). Caubet (1993), following an earlier observation by W. Marçais, mentions only *-Cl*, *-Cr* and *-Cn* as impermissible sequences. On the other hand, McCarthy (2005) makes a more general claim excluding all complex codas with rising sonority (while admitting some exceptions).

(67) Moroccan *CCəC* nouns from Classical Arabic *CVCC* stems:

Moroccan		Middle Arabic	gloss
rʒəl	<	riʒl	‘leg’
bɣəl	<	baɣl	‘mule’
ʔqəl	<	ʔaql	‘mind’
nməl	<	naml	‘ants’
ʔfəl	<	ʔifl	‘child’
ɖhəʔ	<	zuhʔ	‘back’
qʂəʔ	<	qʂr	‘castle’
tmər	<	tamr	‘dates’
ʂmən	<	samn	‘ghee’
ɣʂən	<	ɣuʂn	‘twig’
ʔbəʔ	<	ʔabʔ	‘temper’

For full symmetry, it could be expected that Moroccan *CCəC* nouns, regularly derived from the Classical Arabic *CVVCVC* type, will shift into the other template in cases of heavy violation of onset sonority hierarchy. An example is provided by the noun *wəld* ‘boy’ developed from Classical *walad* contrary to expected *\*wləd*. But such lexical items seem very infrequent<sup>62</sup>, while we do find opposite cases, that is such in which the change of pattern did not occur, although it would be desirable. Consider the collective noun *brək* ‘ducks’ and its singulative form *bərka* ‘duck’, or the masculine noun *ʃwəd* ‘horse’ and its feminine derivative *ʃəwda* ‘mare’. If the base nouns were remodeled according to the *CVCC* template, the hypothetical *\*bərka* and *\*ʃəwd* would not only comply with phonotactic criteria, but would also maintain stem identity with related words, which is the pattern found in some other nouns suffixed with *-a*, e.g. *ləft* ‘turnips’ and *ləfta* ‘turnip’.

To sum up, the syllable structure-driven shift from the *CəCC* type to the *CCəC* type is required, while the reverse repair strategy is not always obligatory. This means that the sonority hierarchy constraints imposed on complex codas are dominant and the respective constraints imposed on onsets seem very weak. It must be stressed out that the change of the noun’s template is a morphophonological process understood as a mapping to an (other) existing template of the same category and not merely a phonological metathesis. If the coda problem were to be solved on purely phonological grounds, one

---

<sup>62</sup> This is my impression after browsing through the dictionary (Sobelman and Harrell 1963). Unlike numerous nouns of the three previous kinds, *wəld* was the only representative of the “fourth” type that caught my eye.

could rather expect a much less costly epenthesis ( $C\text{ə}CC > C\text{ə}CCi$  or  $C\text{ə}CC\text{ə}$ ), but it does not take place, because it would produce a new, non-existing nominal template. On the other hand, the actual morphophonological shifting process does not affect typological relations within the category of “noun”, but only shuffles some of its members into another type. It is also important to note that in the particular case discussed here, the difference between the two types of nouns is only structural and free from any association with particular semantic features, which also facilitates the switch of the template.

Let us now turn to the third person masculine singular form of the past tense (or, perfective aspect) whose template,  $CC\text{ə}C$ , coincides with one of the nominal templates, but never with the other, i.e.  $C\text{ə}CC$ . The Classical Arabic source of this form was  $CaCVC$ -a, which first reduced in Middle Arabic to  $CaCVC$  and then to  $CC\text{ə}C$ , due to the regular processes of syncope and vowel reduction mentioned earlier. While Moroccan nouns, with their two patterns, continued the previous situation of two kinds of nominal templates, verbs also continued the historical form of a unique template in the third person masculine singular. Therefore, verbs in their development never faced the problem of complex codas, which underlies the template-shifting strategy in nouns. And what about the “weak” problem of sonority in complex onsets? Assuming that the  $CC\text{ə}C > C\text{ə}CC$  change of the template did take place in few nouns, we can explain it analogously to the other shift as attraction to an existing pattern. In verbs, such strategy is implausible, because there is not a single lexical item of the category which could provide a pattern. In addition, unlike the very broad category of “noun”, “the third person masculine singular form of the past tense” constitutes an extremely narrow category and as such, it does not welcome type allomorphy. If it were to acquire a completely new type by introducing a  $C\text{ə}CC$  template for some of its members, that would happen at a cost of losing its unique category marker. To conclude, if the onset constraints were undominated in Moroccan phonology, they would have to enforce a change of the verbal template at any cost, but weak as they are, they do not have the power to destroy the stability of this salient verbal category.

In the above argumentation, I have disregarded other members of the verbal past tense paradigm, which are now all shown in (68), together with their Classical Arabic source forms. As demonstrated by this paradigm, the  $CC\text{ə}C$  stem variant appears as a major allomorph, while the  $C\text{ə}CC$  stem (underlined) is found only in the third person singular of the feminine gender and the third person plural. (Note that no such alternation is found in Classical Arabic, in which the past tense/perfective stem of non-derived regular verbs had an invariant  $CaCVC$  template.)

(68) The Moroccan and Classical Arabic past tense paradigm of ‘drink’

	Moroccan		Classical Arabic	gloss
1 sg	ʃrəbt	<	ʃaribtu	‘I drank’
2 sg	ʃrəbti	<	ʃaribta (m), ʃaribtī (f)	‘you drank’
3 sg (m)	ʃrəb	<	ʃariba	‘he drank’
3 sg (f)	<u>ʃrəbət</u>	<	ʃaribat	‘she drank’
1 pl	ʃrəbna	<	ʃaribnā	‘we drank’
2 pl	ʃrəbtu	<	ʃaribtum (m), ʃaribtunna (f)	‘you (pl.) drank’
3 pl	<u>ʃrəbu</u>	<	ʃaribū	‘they drank’

In McCarthy’s (2005) account, the steady template of the third person masculine form correlates with the fact of its being the major allomorph within the paradigm. Leaving aside the mathematical details of the analysis, it predicts that two instances of the minor stem allomorph in a paradigm as above, consisting of seven members, is better than three such instances, which would be the case if *ʃrəb* ‘he drank’ were replaced with the hypothetical variant \**ʃərb*, better adjusted to the syllable structure requirements. Formally, the effect follows from high ranking of intraparadigmatic correspondence constraints, specifically, from their dominance over syllable structure constraints. I have argued above that onset constraints, presumably playing a role here, are rather weak. I will show momentarily that correspondence constraints within this particular paradigm are also rather weak. In fact, there does not seem to be any tendency in Moroccan or other modern Arabic dialects to level the verbal paradigm. Therefore, although McCarthy’s (2005) general theory of Optimal Paradigms in an insightful manner captures the gradient character of leveling, operating in a one-by-one fashion, it is not sufficiently supported by the particular evidence of Moroccan, which, together with many other dialects, provides evidence to the contrary, namely, the stability of allomorphy within the verbal paradigm.

Many modern dialects exhibit stem allomorphy similar to that of Moroccan, in which the forms of the third person singular feminine and the third person plural share the same stem allomorph, which is distinct from that shared by all remaining paradigm members. Representative examples are included in (69), cited from Danecki (1989). Some other dialects have developed as many as three different allomorphs in the past tense of all or at least one type of verbs, cf. the Palestinian verb ‘understand’: *ʃihim* ‘3sg (m)’, *ʃihim-t* ‘1sg’, *ʃihim-at* ‘3sg (f), or Kuwaiti ‘drink’: *ʃirib* ‘3sg (m)’, *ʃiribt* ‘1sg’, *ʃarbat* ‘3sg (f)’.

(69) Past tense allomorphs in Arabic dialects

region	3sg (m) & 1sg	3sg (f) & 3pl	gloss
Kuwait	kitab, kitabt	ktibat, ktibaw	‘write’
Iraq	kitab, kitabt	kitbat, kitbow	‘write’
Syria	daras, darast	darset, darsu	‘study’
Egypt	fihim, fihimt	fihmit, fihmu	‘understand’
Libya	xdem, xdemt	xedmet, xedmū	‘work’
Tunisia	qlib, qlibt	qilbit, qilbu	‘turn’
Mauritania	lbəs, lbəst	ləbsət, ləbsu	‘dress’

In the above data, we can distinguish three different patterns of stem allomorphy sketched in (70) below. In reality, there are even more of them, given the existing three-way distinctions, as well as the fact that one language may combine two patterns depending on phonological properties of the verb. For example, in addition to the alternating pattern, certain dialects (e.g. Egyptian, Palestinian) also exhibit an invariant pattern characterizing verbs with the stem vowel *a*, as *katab* ‘write’ or *daras* ‘study’, etc.

(70) Some of the past tense schemata of stem allomorphy in Arabic dialects

major allomorph	3sg (f) & 3pl	examples of regions
CVCVC	CCVC	Kuwait
CVCVC	CVCC	Iraq, Syria, Egypt
CCVC	CVCC	all Maghrib

In all modern dialects, stem alternation within the verbal paradigm was independently developed in accordance with particular phonological rules of each language. Still, it is conspicuous that no attempt has been made to *suppress* some of them at least in some languages for the purpose of leveling. Technically speaking, it could be said that the constraints which would have to be violated in order to maintain paradigmatic stem identity are just too high to be disobeyed. However, it can be proved at least for certain dialects that this is not true, since in some other environments analogy takes place with simultaneous violation of dominant constraints. I will briefly recall a well-known case of Palestinian and its object clitics (cf. among others, Kenstowicz and Kisseberth 1979, Kenstowicz 1996, Kiparsky 2000).

In Palestinian, syncope affects the unstressed vowel *i* in a (non-final) light syllable. Regular application of syncope results in above-mentioned alternations in the verbal paradigm. However, the rule is systematically suppressed in clitic groups consisting of a verb and an object pronoun. A minimal pair as in (71) illustrates the problem.

(71) Palestinian syncope and its suppression

a/ 1pl past tense verb	b/ 3sg (m) verb + clitic <i>na</i>
<i>fhímna</i> (* <i>fhímna</i> ) 'we understood'	<i>fhímna</i> (* <i>fhímna</i> ) 'he understood us'

All analyses of this issue connect the lack of syncope in (71b) to the form of the independent verb *fhim* 'he understood', in which the first vowel (by being stressed) is not deleted. The relation between the verb and the whole clitic group is maintained either by appealing to two different levels of representation (e.g. through a cyclic rule application as summarized in Kenstowicz and Kisseberth 1979, or through level constraint evaluation, as in Kiparsky 2000), or by parallel evaluation of a clitic group and its Base (as in Kenstowicz 1996). I will assume the latter analysis as formally close to the OT version underlying this work. Interpreted in terms of extended correspondence constraints, it relies on ranking Cor- $\{fhímna: fhim\}$  above the respective "syncope" constraint  $*i]σ$ , which, however, dominates other correspondence constraints involving this verb. As illustrated in (72), a schematic class *A* includes only correspondence constraints between verbs with cliticized object pronouns and respective bare verbs, which constitute their Bases. All other possible pairs of correspondence constraints fall into class *B* and, as dominated by the syncope constraint, are not relevant. They include pairs (in both orders) of all verbal paradigm members, as well as mirror-image correspondence constraints of those included in class *A*, i.e. from bare verbs to clitic groups – their potential Bases. There is one more detail to note in (72), which accounts for the fact of partial (segmental only) identity in class *A* and their difference in stress. I assume that this follows from an undominated constraint family predicting the placement of word stress in Palestinian.

(72) Correspondence ranking for *fhim*-stem family

Stress  $\gg$  Cor-*A*  $\{fhímna$  '3sg (m)+1sg obj.': *fhim* '3sg (m)', etc. $\} \gg *i]σ \gg$   
Cor-*B*  $\{fhim$  '3sg (m)': *fhim-at* '3sg (f), *fhim* '3sg (m)': *fhímna* '3sg (m), etc. $\}$

A question arises: if Palestinian can suppress the syncope constraint and promote Cor-*A* constraints, why not do the same for the intraparadigmatic stem identity and promote the relevant Cor-*B* constraints in a similar fashion? We could imagine, for example, a non-alternating stem *fhim* in the complete paradigm, projected in (73b), as it is in the case of verbs with the vowel *a*, cf. *daras* 'study' in (73c). Or at least a paradigm with an alternation reduced to two allomorphs as in Maghribian dialects, i.e. with the third person singular masculine leveled as  $*fhim$ .

(73) Palestinian past tense paradigm

	a/ actual <i>fihim</i>	b/ hypothetical <i>fihim</i>	c/ actual <i>daras</i>
1sg	fhímt	fihímt	darást
2sg (m)	fhímt	fihímt	darást
2sg (f)	fhímti	fihímti	darásti
3sg (m)	fihim (*fhím)	fihim	dáras
3sg (f)	fihmat	fihimat	dárasat
1pl	fhímna	fihímna	darásna
2pl	fhímtu	fihímtu	darástu
3pl	fihmu	fihimu	dárasu
	‘understand’		‘study’

Similar questions may be posed in reference to other dialects as well. For example in all Maghribian, the third person singular feminine form could easily remodel its CV structure and merge with the first or the second person (cf. a merger of 1sg and 2sg (m) in Palestinian). The fact that none of the above hypothetical analogical developments actually takes place can only mean that there is not much pressure for them to occur. The stability of allomorphy within the Arabic verbal paradigms links to the above-mentioned general thought of substantial semantic distance among verbal paradigm members. But in the particular case here, it is also supported by the strength of templates associated with each individual form. The patterns illustrated in this section characterize the “regular” three-consonantal verbs, which constitute the large majority of all verbs, and which occur in all frequencies, including the highest ones. The template of each form is thus very well entrenched in language usage and there is no need for stem leveling. In fact, this “regular” pattern is much more salient than a number of other minor patterns of verbal inflection existing in Classical Arabic and in a similar fashion in modern dialects, which also exhibit stem allomorphy. For example, the so-called “hollow” verbs are limited to a small group<sup>63</sup> of rather frequent verbs whose stem alternates in the past tense between  $C\bar{a}C$  and  $CuC(CiC)$ , cf. the pairs of 3sg (m) and 1sg: *kān:kunt* ‘be’, *qām:qumt* ‘wake up’, *qāl:qult* ‘speak’, etc. But despite its small size, this class does not seem to be aiming at leveling, either, just as in the case of Polish and English data discussed earlier in this chapter.

---

<sup>63</sup> It is hard for me to estimate their exact number, but presumably they are fewer than twenty.

## CHAPTER 6

# Analogy vis-à-vis Zipf's frequency laws

### 6.1. Zipf's laws – the background

The inverse relation between the frequency of a word (or other linguistic unit) and its length (complexity) has been known since Zipf's (1935) statistical laws. For the purpose of the discussion here it suffices to state these principles informally as: the shorter the word, the higher its frequency, and *vice versa*, longer words have lower frequencies than shorter ones.<sup>64</sup> The easiest way to observe Zipf's laws is to glance at a frequency dictionary of any language to see how consistently the rank of a word inversely correlates with its length, or to count words of different length occurring in a sample text, or to compare forms and frequencies of near synonyms. For example, the British National Corpus shows the following frequencies for the given word-forms of English verbs: *go* (one syllable, two sounds): 25324 times, *walk* (one syllable, three sounds): 1199 times, *stroll* (one syllable, five sounds): 44 times, *saunter* (two syllables, six/five sounds): 6 times, *promenade* (three syllables, eight sounds): no occurrences.<sup>65</sup>

Zipf's laws are not only static generalizations about linguistic facts, but they also have the power to enforce changes in the event of their violation. The dynamic character of Zipf's laws is documented in the vast corpus of literature on grammaticalization. Once a word extends its sphere of use, i.e. it becomes more general in meaning, and consequently, more frequent in usage, its phonological structure considerably diminishes (cf. among others, Bybee 1985, Bybee *et al.* 1994, Heine 1993, Traugott and Heine 1991). However, most studies on grammaticalization treat phonological reduction by way of digression from the semantics which lies in the center of this research.

A quite different perspective is taken by Witold Mańczak, who recognizes this kind of reduction as a historical process of equal importance to regular sound change or analogy. What he terms "irregular phonological development caused by frequency" has been the focus of his research, supported by detailed cross-linguistic evidence (especially from Slavic and Romance) and published in a number of works since early 1960s until most recently (e.g. Mańczak 1965, 1969, 1977, 1988, 1990). Mańczak's analyses often coincide with the conclusions drawn from the research on grammaticalization, since higher frequency and grammaticalization typically go together. And just as grammaticalization is more and more often claimed to be unidirectional and universal, Mańczak, too, argues that his "irregular development" appears very "regular". The

---

<sup>64</sup> The best known among Zipf's laws (often referred to in the literature just as "Zipf's law" in the singular) states that the frequency of the word is inversely proportional to its rank. But this law will not be relevant to the discussion in this chapter.

<sup>65</sup> It is interesting to observe how the speakers intuitively feel the "appropriate" or "inappropriate" length of a given word or phrase. Let me illustrate this with a real case. A Polish speaker, who does not know any French, once told me how surprised he was to find out that *quelque chose* means 'something' (in Polish *coś*). During a long car-trip, he was listening to two people speaking French and was struck by the frequent occurrence of this sequence of sounds. 'What does that *quelque chose* that you are repeating all the time, mean?' – he asked them. Once given the answer, he was so amazed that found this fact worth enough to discuss with me on some occasion. His exact words were: 'I could not believe that such a "nothing" was so long!' Indeed, French trisyllabic *quelque chose* seems to be an exception with respect to Zipf's laws.

example in (72), adopted from Mańczak (1977: 177) illustrates a common development of conjunctions from adverbs, accompanied by the reduction of structure. Because in these particular languages, the source adverbs are still in use, we can compare their “regular” development and the “irregular” development of the more frequent conjunctions. (Note that in Spanish both words are pronounced identically, but the adverb has still more “structure” in orthography.)

(72) Regular *versus* irregular development

	adverb	conjunction
Polish	<i>więcej</i> ‘more’	<i>więc</i> ‘so’
Italian	<i>mai</i> ‘never’	<i>ma</i> ‘but’
Provençal	<i>mais</i> ‘more’	<i>mas</i> ‘but’
Catalan	<i>may</i> ‘never’	<i>mes</i> ‘but’
Portuguese	<i>mais</i> ‘more’	<i>mes</i> ‘but’
Spanish	<i>más</i> ‘more’	<i>mas</i> ‘but’

But changes in frequency of use do not always correlate with the degree of grammaticalization, or at least not so evidently. Some cases of allomorphy provide the best example of the same meaning realized in a smaller/larger form, depending on frequency. Let us consider another example from Mańczak (1977: 192), concerning the development of the Slavic preposition \* *kъ*, which in Old Polish alternated between *ku* and *k*. According to him and contrary to the common opinion, *ku* was an older form, which shortened to *k* in frequently used phrases, e.g. with pronouns, cf. *k niemu* ‘to, towards him’ vs. *ku grobu* ‘towards the grave’. In this context, it is worthwhile to add to Mańczak’s observations and note that this particular preposition became very rare in modern Polish, eliminated by *do* ‘to’ and other prepositions. But when used now, it may have only the fuller form *ku*, hence in the contemporary language: *ku niemu* ‘to, towards him’ and *ku grobowi* ‘towards the grave’. The modern form, which apparently contradicts the unidirectionality principle in grammaticalization, nicely complies with two facts: 1/ elimination of allomorphy in rare words, and 2/ adjustment of the size of a linguistic unit to its frequency.

Synchronic grammars of various languages offer a plentiful supply of allomorphy similar to the Old Polish case. The following section presents some examples from Standard Swahili.

## 6.2. Zipf’s laws in Swahili morphophonology

The negative prefix *ha* is followed in verbs by a personal or a noun class prefix. It keeps its full form before all noun class markers, even vowel-initial, but is reduced to *h* before vowel-initial prefixes of 3<sup>rd</sup> (*a*) and 2<sup>nd</sup> (*u*) persons singular. The latter case is especially interesting because the *u* is homonymic with the marker of classes 3 and 11, in which the reduction does not take place. As a result, negated verbs in cl. 3/11 are always one syllable longer than the corresponding 2<sup>nd</sup> person verbs, cf. (73).

(73) Irregular shortening in the 2<sup>nd</sup> person

a/	2 <sup>nd</sup> sg	/ha+u/ > <i>hu</i>	e.g. <i>hukuanguka</i> ‘you (sg) did not fall’
b/	cl. 3	/ha+u/ > <i>hau</i>	e.g. ( <i>mti</i> ) <i>haukuanguka</i> ‘(tree) did not fall’
	cl. 11	/ha+u/ > <i>hau</i>	e.g. ( <i>ushanga</i> ) <i>haukuanguka</i> ‘(bead) did not fall’

Irregular shortening also takes place in the 1<sup>st</sup> person singular verbs, as demonstrated in (74a). Unlike (73a), which resembles other morphophonemic processes in Swahili involving vowel coalescence, the reduction in (74a) is purely idiosyncratic and does not have a parallel anywhere else in the grammar of this language. Let us also observe that no reduction takes place in plural personal prefixes, cf. (74d-f). We can see in the complete paradigm of all persons/numbers in (74) that, due to the irregular shortening, all singular forms are one syllable shorter than all plural forms, which in turn are just as long as verbs with noun class prefixes. This result is clearly frequency-driven, since verbs are more often used with personal pronouns than with non-personal pronouns (i.e. noun class markers), and verbs in the singular are more used than in the plural.

(74) Irregular shortening in personal prefixes

a/	1 <sup>st</sup> sg	/ha+ni/ > <i>si</i>	e.g. <i>sikuanguka</i> ‘I did not fall’
b/	2 <sup>nd</sup> sg	/ha+u/ > <i>hu</i>	e.g. <i>hukuanguka</i> ‘you (sg) did not fall’
c/	3 <sup>rd</sup> sg	/ha+a/ > <i>ha</i>	e.g. <i>hakuanguka</i> ‘(s)he did not fall’
d/	1 <sup>st</sup> pl	/ha+tu/ > <i>hatu</i>	e.g. <i>hatukuanguka</i> ‘we did not fall’
e/	2 <sup>nd</sup> pl	/ha+m̩/ > <i>ham̩</i>	e.g. <i>ham̩kuanguka</i> ‘you (pl) did not fall’
f/	3 <sup>rd</sup> pl	/ha+wa/> <i>hawa</i>	e.g. <i>hawakuanguka</i> ‘they did not fall’

Another case of allomorphy in Swahili concerns the infinitive (or, class 15 “gerund”) prefix, always realized as *ku* before a consonant-initial stem, and occasionally alternating as *ku~kw* before a vowel-initial stem. In this environment, the *ku* variant is unmarked and *kw* appears only with the stems of highest frequency, which are often, but not always grammaticalized. I illustrate the issue with some examples in (75). The two columns of figures indicate the frequencies of given forms as they appear in the sample corpora of two different styles, the newspaper *Nipashe* and a collection of literary works (*Books*) of the Helsinki Corpus of Swahili. The occurrences of the other variant of the prefix are listed in parentheses.

The most frequent lexical item of this group is the complementizer *kwamba* ‘that’, as in ‘he said *that*’, cf. (75a), etymologically resulting from the verb ‘to say’. Although the base verb is not used any more, its derived forms are, with the regular *ku* in the infinitive, cf. *kuambiana* ‘to say to each other’. The case of *kuanza~kwanza* in (75b-c) demonstrates that if the infinitive lexical source is still in usage alongside the grammaticalized form, the former, as less frequent, keeps the *ku* allomorph, while the latter has the *kw* allomorph. Notice that in this particular example, the frequency of the verb is also very high, which is due to its ‘second cycle’ of grammaticalization into the

inchoative aspect.<sup>66</sup> The verb *kwenda* ‘to go’ in (75d), as the most common verb of motion has very high frequency, which makes it occur almost always with the *kw* prefix (but note 10 occurrences of *ku* in *Books*). Let us observe that its much rarer derivative in (75e) appears only with the *ku* allomorph. The final example in (75f) provides additional evidence that it is frequency and not grammaticalization *per se* which triggers the reduction of the prefix. The infinitive *kuelekea* “to go towards” is highly grammaticalized into the meaning ‘towards, (in) the direction of’, as illustrated by the corpus examples in (76), in which the original meaning of motion is no longer present. But this fact, together with low frequency, does not trigger the prefix’s reduction. The verb *kuelekea* sharply contrasts with *kwenda*, which undergoes the reduction, although it is only slightly grammaticalized into the meaning ‘to, in the direction’, when used with another verb of motion, e.g. *ameondoka nyumbani kwenda shuleni* ‘(s)he left home to go to school’.

(75) Frequency-driven *ku* > *kw*

		<i>Nipashe</i>	<i>Books</i>
a/	<i>kwamba</i> ‘that’	7209	3287
b/	<i>kuanza</i> ‘to begin’	751	194
c/	<i>kwanza</i> ‘first, firstly’	1064	1429
d/	<i>kwenda</i> ‘to go’	1536 ( <i>ku</i> -0)	724 ( <i>ku</i> -10)
e/	<i>kuendelea</i> ‘to continue’	719 ( <i>kw</i> -0)	167 ( <i>ku</i> -0)
f/	<i>kuelekea</i> ‘to go towards; towards’	198 ( <i>kw</i> -0)	186 ( <i>kw</i> -0)

(76) Grammaticalization stages of *kuelekea*

- a/ Lakini polepole mawazo ya kisiasa nchini yakabadilika na kuelekea mashariki.  
‘But slowly the political views in the country changed and directed to the East.’
- b/ Picha kubwa mbili [...] zimetundikwa ukutani kuelekea watazamaji.  
‘Two big pictures [...] are hung on the wall facing the viewers.’
- c/ Kuelekea kaskazini ya sebule kuliwekwa viti vinne [...].  
‘Towards the North of the lounge, four chairs were placed [...].’

The next example involves allomorphy of the future marker *ta~taka*. The latter variant is used only in relative clauses of the “synthetic” type, i.e. such in which a relative pronoun (italicized in (77)) is infixed between the tense marker and the verb stem. Historically, *taka* constitutes an older form, which preserves the complete stem of its etymological source, the verb ‘want’ (cf. Heine 1993). This stem was shortened to monosyllabic *ta* in more frequent forms without the relative pronoun. No comparable allomorphy is found in other tenses, the past tense with the *li* marker and the present tense with the *na* marker, which are monosyllabic in both environments. This points out that there is no special prosodic structure requirement imposed on the relative clause. In (77), I list frequencies of sample common verbs in 3<sup>rd</sup> person singular for all three tenses as they occur in the *Nipashe* part of the Helsinki corpus. The four columns in (77a) indicate from left to right: the form without a relative pronoun, with a subject relative

<sup>66</sup> Such multiple „cycles” of grammaticalization are nothing unusual in Swahili. For example, the verb *taka* ‘want’, discussed later in this section, grammaticalized into the future marker *ta* and is currently undergoing slight grammaticalization into an auxiliary indicating that something may happen soon or almost happened, cf. Heine (1997).

pronoun, with an object relative pronoun (appropriate to the most commonly designated object of the verb) and with the commonly used relative pronoun *po*, which in most cases refers to time (and sometimes to place). The examples in (77b-d) are analogous with somewhat abbreviated notation. The bolded figures indicate the occurrences of the verb without a relative pronoun followed by the total sum of occurrences of the verb with all possible (not only the exemplified) pronouns. Other figures indicate occurrences of the verb with the most common sample pronouns. Two obvious conclusions can be drawn from these data. First, the verbs with relative pronouns are less frequent than the verbs without them (but note an exceptional case in (77b)). Secondly, the future tense is used less than other tenses and in particular, occurrences of future verbs with relative pronouns are extremely infrequent. The use of the disyllabic allomorph of the future marker in this particular environment is thus motivated by Zipf's frequency laws. Notice, however, that the *taka* allomorph, although strengthened by its open-class template, is not very salient with respect to token frequency. Therefore, it is under the pressure to undergo the leveling. Indeed, in other Swahili dialects such leveling took place and a monosyllabic future marker occurs in both environments.

(77) *Nipashe* frequencies of verbs without and with a relative pronoun

a/ alikwenda: <b>136:56</b>	aliyekwenda 10	alikorwenda 10	alipokwenda 36
'(s)he went'	'who went'	'where (s)he went'	'when (s)he went'
anakwenda <b>51:11</b>	anayekwenda 5	anakorwenda 2	anapokwenda 4
'(s)he goes'	'who goes'	where (s)he goes'	'when (s)he goes'
atakwenda <b>11:2</b>	atakayekwenda 2	atakakorwenda 0	atakapokwenda 0
'(s)he will go'	'who will go'	'where (s)he'll go'	'when (s)he'll go'
b/ alikuja <b>25:8</b>	<i>ye</i> 4: <i>ko</i> 0: <i>po</i> 4		
'(s)he came'	'who came': 'where (s)he came': 'when (s)he came'		
anakuja <b>8:5</b>	<i>ye</i> 4: <i>ko</i> 0: <i>po</i> 1		
'(s)he comes'	'who comes': 'where (s)he comes': 'when (s)he comes'		
atakuja <b>4:6</b>	<i>ye</i> 5: <i>ko</i> 0: <i>po</i> 1		
'(s)he will come'	'who will come': 'where (s)he will come': 'when (s)he will come'		
c/ alifanya <b>121:104</b>	<i>ye</i> 17: <i>yo</i> 5: <i>po</i> 13		
'(s)he did'	'who did': '(work) which (s)he did': 'when (s)he did'		
anafanya <b>40:23</b>	<i>ye</i> 9: <i>yo</i> 5: <i>po</i> : 2		
'(s)he does'	'who does': '(work) which (s)he does': 'when (s)he does'		
atafanya <b>34:12</b>	<i>ye</i> 5: <i>yo</i> 1: <i>po</i> : 2		
'(s)he will do'	'who will do': '(work) which (s)he will do': 'when (s)he will do'		
d/ alisema <b>27977:65</b>	<i>ye</i> 10: <i>yo</i> 12: <i>po</i> 7		
'(s)he said'	'who said': '(words) which (s)he said': 'when		
anasema <b>52:9</b>	<i>ye</i> 2: <i>yo</i> 0: <i>po</i> 2		
'(s)he says'	'who says': '(words) which (s)he says': 'when		
atasema <b>1:0</b>	<i>ye</i> 0: <i>yo</i> 0: <i>po</i> 0		
'(s)he will say'	'who will say': '(words) which (s)he will say': 'when		

Zipf's laws are important enough to find a place in the architecture of OT. In a sense, they are already there as "minimality" constraints, which impose a minimal length requirement (e.g. disyllabicity) on lexical words (cf. e.g. McCarthy and Prince 1990, 1994). Mirror-image "maximality" constraints, requiring the structure be no larger than a given prosodic unit (e.g. a foot, a syllable) have been introduced in OT mostly in the context of reduplication (cf. McCarthy 1995). Constraints which, as in the above examples, are sensitive to particular lexical contexts can be analyzed in OT by means of highly ranked specific constraints (also named "lexical" or "parochial", cf. among others, Hammond 1999, Kraska-Szlenk 1997, 1999b, 2003, Myers 1999, Pater 2000, to appear, Russell 1995). But, to my knowledge, neither "minimality" or "maximality" constraints have been discussed in OT from the perspective of language use and Zipf's laws, which provide a motivation for them.

Some of such "proportionality" constraints (collapsing "minimality" and "maximality" under one term) often have a general formulation in that they affect a given class, and not as in the earlier examples, only particular members of it. For example, the minimal size requirement of disyllabicity in Swahili refers to all independent words (nouns, verbs, etc. including pronouns), without exempting from it the most frequent ones. This effect is formally phrased by a pair of non-violable constraints: LexWord $\theta$ Foot ("a lexical word must contain a foot") and FootBin ("a metrical foot must be binary"). A number of various strategies guarantee that minimality be obeyed on the surface, at the cost of faithfulness violations, as exemplified in (78) below (cf. Batibo and Rottland 1992 for a full account).

(78) Swahili "minimality" strategies

word	regular morphophonology	"minimality" morphophonology
nominal cl. 1/3	non-syllabic prefix <i>m</i> , e.g. <i>mtoto</i> 'child', <i>mtama</i> 'sorghum'	syllabic prefix <i>m̩</i> e.g. <i>m̩tu</i> 'person', <i>m̩ti</i> 'tree'
nominal cl. 9/10	homorganic prefix <sup>N</sup> , e.g. <sup>ɔ</sup> <i>goma</i> 'drum', <sup>m</sup> <i>buzi</i> "goat"	homorganic syllabic prefix <sup>N̩</sup> e.g. <sup>ɲ</sup> <i>ge</i> 'scorpion(s)', <sup>m̩</sup> <i>bu</i> 'mosquito(s)'
nominal cl. 9/10	no prefix, e.g. <i>pana</i> 'broad-cl. 9/10', <i>tai</i> 'eagle'	homorganic syllabic prefix <sup>N̩</sup> , e.g. <sup>m̩</sup> <i>pya</i> 'new-cl. 9/10', <sup>n̩</sup> <i>ta</i> 'wax'
nominal cl. 5	no prefix, e.g. <i>pana</i> 'broad-cl. 5', <i>wazo</i> 'thought' (pl. <i>mawazo</i> )	prefix <i>ji</i> e.g. <i>jipya</i> 'new-cl. 5', <i>jiwe</i> 'stone' (pl. <i>mawe</i> )
cl.10 pl. of cl. 11	no prefix, e.g. <i>shanga</i> 'beads' (sg. <i>ushanga</i> ), <i>kuta</i> 'walls' (sg. <i>ukuta</i> )	prefix <i>ny+u</i> , e.g. <i>nyusi</i> 'eye-brows' (sg. <i>usi</i> ), <i>nyuso</i> 'faces' (sg. <i>uso</i> )
impera- tive	no prefix, e.g. <i>soma</i> 'study-sg.', <i>lala</i> 'sleep-sg.'	prefix <i>ku</i> , e.g. <i>kula</i> 'eat-sg.', <i>kunywa</i> 'drink-sg.'
personal pronoun	monosyllabic, when encliticized e.g. <i>nami</i> 'and/with me', <i>nawe</i> 'and/with you', <i>nasi</i> 'and/with us'	reduplicated, when independent, e.g. <i>mimi</i> 'I, me', <i>wewe</i> 'you', <i>sisi</i> 'we'

Similarly, morphophonological alternations triggered by general application of maximality constraints are attested in Swahili, too. For example, a class of concordial prefixes of the so-called pronominal series have allomorphs of the shape  $CV\sim C(G)$ . The  $CV$ -allomorph occurs before a consonant-initial stem and the  $C(G)$ -allomorph before a vowel-initial stem. As a result, there is never a hiatus or consonantal cluster at the prefix-stem juncture. Similarly, other prefixes alternate as  $V\sim G$ . The full set of alternating prefixes with relevant examples is included in (79-80). In one case of the non-syllabic version of the prefix (cl. 7), the glide is fused with the preceding consonant causing its palatalization.

(79) Swahili “pronominal” prefix alternations:

$u\sim w$ (cl. 3/11)	$i\sim y$ (cl. 4/9)	$wa\sim w$ (cl. 2)	$ya\sim y$ (cl. 6)
$li\sim l$ (cl. 5)	$zi\sim z$ (cl. 10)	$ki\sim ch$ (cl. 7)	$vi\sim vy$ (cl. 8)
$ku\sim kw$ (cl. 15/17)	$pa\sim p$ (cl. 16)	$ṃ(mu)\sim mw$ (cl. 18)	

(80a) Syllabic prefix in a C-initial context:

- tense marker *na, li, me, ta* etc., e.g. *tunda limeiva* ‘the fruit is ripe’, *habari zitafika* ‘the news will arrive’, *vitu vilipotea* ‘things got lost’;
- copula *na*, e.g. *chanzo kina mwisho* ‘the beginning has an end’, *mti una majani* ‘the tree has got leaves’;
- locative copulas *po, ko, mo*, e.g. *kalamu ipo mezani* ‘the pen is on the table’, *watu wako nyumbani* ‘the people are in the house’;
- demonstrative *le*, e.g. *kitu kile* ‘that thing’, *vitu vile* ‘those things’, *jiwe lile* ‘that stone’, *mawe yale* ‘those stones’;

(80b) Non-syllabic prefix in a V-initial context:

- tense marker *a*, e.g. *gari laibwa* ‘the car is stolen’, *habari zafika*, ‘the news are coming’, *UKIMWI waua* ‘AIDS kills’;
- associative particle *a*, e.g. *miti ya machungwa* ‘orange trees’, *vitabu vya Ali* ‘Ali’s books’, *nyumba za kupanga* ‘houses for rent’;
- possessive pronouns, e.g. *moyo wangu* ‘my heart’, *moyoni mwangu* ‘in my heart’, *nyumba yao* ‘their house’, *vitu vyetu* ‘our things’, *somo lako* ‘your class’;
- relative pronoun *o*, e.g. *kitabu nilichokinunua* ‘the book which I bought’, *siku zijazo* ‘coming days’, *miezi iliyopita* ‘past months’;<sup>67</sup>
- pronoun *ote*, e.g. *habari zote* ‘all news’, *mambo yote* ‘all matters’, *watu wote* ‘all people’, *vitu vyote* ‘all things’;
- pronoun *enye*, e.g. *mti wenye matunda* ‘fruit tree’, *jambo lenye maana* ‘important matter’, *chumba chenye bafu* ‘room with a bathroom’;

<sup>67</sup> The glide *w* is deleted in this context, e.g. *mwezi uliopita* (*\*uliwopita*) ‘past month’, *watu waliokuja* (*\*waliwokuja*) ‘people who came’.

A possible analysis of the prefixes allomorphy may appeal to the Onset constraint (“all syllables must have onsets”) and the effect of “emergence of the unmarked”, i.e. a situation in which particular morphemes are subject to markedness constraints normally disobeyed in a given language (cf. McCarthy and Prince 1994, Kraska-Szlenk 1999b). In Swahili, Onset is frequently violated at morpheme boundaries, as well as morpheme-internally, cf. *maua* (*ma+ua*) ‘flowers’, *tuendele* (*tu+end+el+e+e*) ‘let’s continue’, *aoe* (*a+o+e*) ‘let him get married’, etc. This points out that Onset has low ranking, below the faithfulness constraints. But it can be assumed that in the case of concordial prefixes, Onset dominates respective faithfulness constraints, enforcing idiosyncratic vowel elision or gliding. The ranking: Faith-Roots etc. >> Onset >> Faith-Prefix<sub>[concord]</sub> predicts that while hiatus will remain on the surface in various contexts, it will be resolved at the cost of the loss of prefix structure, e.g. *li+a>la* (\**lia*). While such an analysis may seem quite appropriate for these particular data, it will be untenable for other similar facts of Swahili to which I will come later in this section. Therefore, it is worthwhile to consider an alternative account in the vein of Zipf’s laws, involving a maximality constraint which will give preference to non-syllabic realization of this set of prefixes.

A specific constraint No-Syll{*u,i,li,zi* etc.}, which reads “concord prefixes {*u,i,li,zi* etc.} must not contain a syllable”, dominating Max{*u,i,li,zi* etc.} will correctly predict that a prefix’s vowel is glided/deleted, where it may without markedness violations. For the sake of space, I ignore some details of analysis involving markedness and general faithfulness constraints (preventing the deletion of the stem consonant following the prefix), which together guarantee that as much of the prefixes’ structure is saved on the surface as possible, provided No-Syll is satisfied. In addition, a dominant constraint against complete deletion of a morpheme (No-Zero-Morph, cf. Kraska-Szlenk 1999b) may be required to enforce the realization of the prefixal vowel in the cases where a *C(G)* variant is impossible for phonotactic reasons. Hence, the ultimate ranking: No-Zero-Morph >> No-Syll{*u,i,li,zi* etc.} >> Max{*u,i,li,zi* etc.} will basically account for the two variants of the prefixes in the two different phonological environments.

A more exact account of these data should include an additional explanation. Observe that all contexts of the reduction of the prefixes in (80b) involve their occurrence before either a fully grammatical affix (as the first four) or a highly grammaticalized lexical morpheme (as the last two). This does not seem to be coincidental, since a prefix does not shorten before a vowel-initial lexical root. The relevant environment is found when a prefix occurs in a subject or an object position before a verb. As exemplified in (81a-b), various kinds of a vocalic hiatus result on the surface, although they could be easily resolved by gliding/vowel deletion similarly to the cases in (80b). The difference in behavior between (80) and (81) can be explained in terms of frequency, since all combinations of the prefix and a following morpheme in (80b), as highly grammaticalized, are more frequent than a combination of a subject marker and a given verb in the subjunctive form (which is the only context of the subject prefix occurring directly before a verb), or a given verb and an object marker. This reasoning is somewhat supported by a case in (81c), in which the subject marker *tu* of the 1<sup>st</sup> person plural is reduced only in the very frequent expression with the earlier mentioned verb *enda* ‘go’ (in the hortative meaning ‘let’s go’) and not with other vowel-initial verbs in the subjunctive form, which is reminiscent of the earlier discussion of this verb in the context of the infinitive prefix.

(81a) Syllabic subject marker before a V-initial verb:

- cl. 2 *waende* ‘let them go’
- cl. 3 (*mkataba*) *uandikwe* ‘let (the agreement) be written’
- cl. 4 (*miti*) *iangushwe* ‘let (the trees) be felled down’
- cl. 5 (*gari*) *liendeshwe* ‘let (the car) be driven’
- cl. 6 (*maswali*) *yaulizwe* ‘let (the questions) be asked’
- cl. 7 (*kiti*) *kiingizwe* ‘let (the chair) be brought in’
- cl. 8 (*viatu*) *vioshwe* ‘let (the shoes) be cleaned’

(81b) Syllabic object marker before a V-initial verb:

- cl. 2 *nawaomba* ‘I ask them’
- cl. 3 *nauandika* (*mkataba*) ‘I write it (the agreement)’
- cl. 4 *naiangusha* (*miti*) ‘I make them (the trees) fall down’
- cl. 5 *naliendesha* (*gari*) ‘I drive it (the car)’
- cl. 6 *nayauliza* (*maswali*) ‘I ask them (the questions)’
- cl. 7 *nakiingiza* (*kiti*) ‘I am bringing it (the chair) in’
- cl. 8 *naviosha* (*viatu*) ‘I am cleaning them (the shoes)’

(81c) Syllabic versus non-syllabic subject marker *tu*:

- |                                   |                                 |
|-----------------------------------|---------------------------------|
| <i>tuimbe</i> ‘let’s sing’        | <i>tuingie</i> ‘let’s enter’    |
| <i>tuongoze</i> ‘let’s lead’      | <i>tuulize</i> ‘let’s ask’      |
| <i>tuendelee</i> ‘let’s continue’ | <u><i>twende</i></u> ‘let’s go’ |

An OT analysis of the above data may appeal either to a “special” property of the prefix, which remains syllabic in spite of the phonological environment conditioning its reduction, or to a special syllabification property of the verbal root category. The former strategy will have to restrict the specific No-Syll constraint by subcategorizing it to the “grammatical”, open class environments of (80b). As mentioned earlier, such an account reflects a maximality constraint operating in frequent contexts and as such, it makes sense. A second analysis would involve a dominant alignment constraint requiring that the left edge of the verbal root must coincide with the left edge of the syllable (Align-Syll-L<sub>[verb]</sub>). Naturally, deletion of the prefixal vowel induces a violation of the alignment constraint by introducing a “foreign” onset sponsored by a different morpheme. From the perspective of language use, Align-Syll-L<sub>[verb]</sub> also makes sense, since it increases the salience of the lexical root and is functionally very similar to the prosodic minimality constraints mentioned earlier.

In contrast to the “pronominal” series of prefixes described above, whose reduction is to a large extent phonologically conditioned, the “nominal” series of prefixes exhibit peculiar behaviors not always justifiable in phonological terms. The most puzzling cases are those in which the choice of a given allomorph produces maximally marked structures, while there exists a less marked allomorph as a candidate. As an illustration, I will use homophonous prefixes of noun classes 1 and 3, which have the following allomorphs: *m*, *mu*, *mw* and *m*. The first allomorph, *m*, occurs only before a

monosyllabic root and is triggered by the minimality constraint mentioned earlier. The second allomorph, *mu*, is found before stems starting with *u* or *h* (sometimes in variation with *m* or *mw*), before other consonants (mostly *s* or *sh*) in a very small group of loanwords (with variation, too) and before a few monosyllabic roots. The allomorph *mw* occurs before all other vowel-initial stems and the allomorph *m* before all other consonant-initial stems (disyllabic or longer). The examples of both noun classes are shown in (82), where variation is indicated whenever it is marked as possible in Miachina's (1987) comprehensive Swahili-Russian dictionary. What is interesting about these data is the existence of numerous nouns, as those shown in (82b), in which sonority hierarchy is heavily violated in the initial syllable. If markedness constraints played any role in allomorph selection, obviously *mu* would be a better choice. But for almost all consonant-initial nouns (with the few exceptions as in (82a)), it is not even a free variant option. Therefore, the Swahili grammar must prefer the *m* allomorph on some other grounds. It is my contention that this preference relates to the shortness of the prefix, similarly to the cases discussed previously. Let us observe that the specific No-Syll constraint imposed on the *mu* prefix must be ranked relatively high to be able to overrule the \*NC markedness constraint.

(82a) Nouns prefixed with *m̥*, *mu* or *mw* (including alternating *m*):

class 1:

*m̥tu* 'person'  
*m̥fu* 'dead person'  
*muumba~mwumba* 'creator'  
*muuguzi~mwuguzi* 'nurse'  
*mwana* 'child'  
*mwehu* 'lunatic'  
*mwindaji* 'hunter'  
*Muhindi~Mhindi* 'Indian'  
*mhenga* 'ancestor'  
*mhandisi* 'engineer' (Ar.)  
*mshitiri~mushtiri* 'buyer' (Ar.)

class 3:

*m̥ti* 'tree'  
*muwa* 'sugarcane'  
*muungu~mungu* 'god'  
*muujiza~mwujiza* 'miracle'  
*mwaka* 'year'  
*mwezi* 'moon'  
*mwiko* 'taboo'  
*muhindi~mhindi* 'maize'  
*muhanga~mhanga* 'sacrifice'  
*muhula* 'semester' (Ar.)  
*musuli~msuli* 'muscle' (Eng.)

(82b) Nouns with the non-alternating prefix *m*:

class 1:

*mstadi* 'expert'  
*msemaji* 'speaker'  
*mfanyakazi* 'worker'  
*mtemi* 'chief'  
*mkazi* 'inhabitant'  
*mpagazi* 'porter'  
*mzazi* 'parent'

class 3:

*mstari* 'line' (Ar.)  
*msalaba* 'cross' (Ar.)  
*mfuko* 'bag'  
*mtego* 'trap'  
*mkeka* 'mat'  
*mpaka* 'border'  
*mzigo* 'burden'

*mvuvi* ‘fisherman’  
*mdudu* ‘insect’  
*mgeni* ‘foreigner’  
*mbuni* ‘author’

*mvuke* ‘steam’  
*mdomo* ‘mouth’  
*mguu* ‘leg’  
*mbuni* ‘coffee tree’

In most noun classes, prefixes of the “nominal” series have the same form in all nominals, i.e. nouns, adjectives and numerals. But in some cases we can observe differences similar to those distinguished earlier for the “pronominal” series. For example, in the prevocalic environment, the prefix of class 4 has the form *mi* before nominal stems, while it is usually shortened to *my/m* before adjectival stems. As illustrated in (83), the reduction always takes place before *e* (several adjectives), *i* (only two adjectives), but is not very regular before *o*.<sup>68</sup> The interpretation of these data is analogous to the earlier cases of (80b) versus (81): the shorter, non-syllabic prefix occurs with an open class of adjectives, while the longer, syllabic allomorph with a closed class of nouns. It should be noted that the category of “pure” adjectives<sup>69</sup> in Swahili is very small and various attributive notions are expressed by means of verbs or nominal constructions. At the same time, adjectives are typically highly polysemous (as noticeable in some of the translations in (83)) and, consequently, very frequent.

(83) Noun-adjective prefix asymmetries

nouns of class 4:

*miezi* ‘months’  
*miendo* ‘movements’  
*miiko* ‘taboos’  
*miiba* ‘thorns’  
*mioto* ‘fires’  
*mionzi* ‘rays’

adjectives of class 4:

*myeupe* ‘white; plain; empty’  
*myepesi* ‘quick; light; thin’  
*mingi* ‘many’  
*mingine* ‘other’  
*myororo~miororo* (rarer) ‘smooth, soft’  
*miovu* ‘bad’ (?*myovu*)

While all examples in this section involve allomorphy of grammatical morphemes, the following section will focus on allomorphy of a lexical stem. I will argue that in this domain, too, Zipf’s laws can trigger the choice of a lexical allomorph, leading to the partial leveling within a verbal paradigm. The data will be drawn from standard and colloquial Polish.

<sup>68</sup> It is likely that this difference is not conditioned by the vowel quality *per se*, but is due to the frequency, since *e*- and *i*-initial adjectives are more frequent than those starting with *o*.

<sup>69</sup> What I mean by that is the native Bantu adjectives which are used with agreement prefixes as opposed to borrowed adjectives, having an invariant, prefixless form.

### 6.3. Leveling in a verbal paradigm conditioned by Zipf's laws

#### 6.3.1. The paradigm of the past tense of *iść* in standard Polish

The past tense paradigm of the Polish verb *iść* 'walk, go'<sup>70</sup> is exceptional in two ways – it is based on a suppletive stem, which, moreover, alternates in an idiosyncratic fashion. As illustrated in (84), all singular feminine and neuter forms use the stem allomorph [ʃ], while all singular masculine forms are formed with the allomorph [ʃed]. The suffix [w] follows the stem to form the etymological participle *szedł* [ʃedw]. The feminine and neuter participles are additionally followed by the *-a* and *-o* endings, respectively, to become *szła* [ʃwa] and *szło* [ʃwɔ]. The alternation is historically motivated by the vocalization/deletion of the root yer vowel *ɨ* (cf. chapter 4). In the masculine form, the yer vocalized as *e* before the deleted final yer of the inflection, cf. *\*ʃɨdwɨ* > *ʃedw*. In feminine and neuter forms suffixed with the full vowel, the deletion of the yer took place with the subsequent cluster simplification, cf. *\*ʃɨdwa* > *ʃɨwa* > *ʃwa*.

The original participles are used as synchronic third person verbs of the past tense. The masculine form has an optional, less formal pronunciation without the final [w], i.e. [ʃet] with the effect of final devoicing, or, [ʃed] in the context of voicing assimilation. In the first and second persons, the participle occurs with a clitic-like element (for ease of exposition separated in (84) by a hyphen). The plural forms are based exclusively on the stem allomorph [ʃ] which makes their structure appear quite regular (cf. the right-hand columns in (84) representing a paradigm of another verb). The plural participles are formed with the complex suffix *-li* for the masculine gender and *-ły* [wɨ] for the feminine/neuter gender. As in the singular, enclitics are added in first and second persons. The verb *robić* 'do' in the right-hand column of (84) represents a typical Polish verb in the past tense paradigm. Note that in this case the only difference between the masculine and feminine forms of the singular is the vowel of the ending: *e* in masculine and *a* in feminine. This pattern is found in all other verbs and provides a strong basis for template analogy.

---

<sup>70</sup> I give these two translations, because on the one hand, *iść* refers to the motion of humans and animals, but not cars, trains etc., hence 'walk' is more appropriate. On the other hand, the verb is very common, highly grammaticalized and used in many abstract contexts, which is better rendered in English by 'go'.

(84) The past tense paradigms of *iść* ‘go’ and *robić* ‘do’

	singular	plural		singular	plural
1 m	szedł-em [ʃɛdwɛm]	szli-śmy [ʃliɕmɨ]	cf.	robił-em [rɔbiwɛm]	robili-śmy [rɔbiliɕmɨ]
1 f	szła-m [ʃwam]	szły-śmy [ʃwiɕmɨ]		robiła-m [rɔbiwam]	robiły-śmy [rɔbiwiɕmɨ]
2 m	szedł-eś [ʃɛdwɛɕ]	szli-ście [ʃliɕtɕɛ]		robił-eś [rɔbiwɛɕ]	robili-ście [rɔbiliɕtɕɛ]
2 f	szła-ś [ʃwacɕ]	szły-ście [ʃwiɕtɕɛ]		robiła-ś [rɔbiwacɕ]	robiły-ście [rɔbiwiɕtɕɛ]
3 m	szedł [ʃɛdw]	szli [ʃli]		robił [rɔbiw]	robili [rɔbili]
3 f	szła [ʃwa]	szły [ʃwi]		robiła [rɔbiwa]	robiły [rɔbiwi]
3 n	szło [ʃwɔ]	szły [ʃwi]		robiło [rɔbiwɔ]	robiły [rɔbiwi]
	‘I went’ etc.	‘we went’ etc.		‘I did’ etc.	‘we did’ etc.

The allomorphy illustrated in (84) characterizes this one stem only, but is present in all commonly used derivatives of the verb *iść* formed with many different prefixes, which modify the meaning, cf. masculine forms *poszedłem* [pɔʃɛdwɛm] ‘I went (away)’, *wyszedłem* [viʃɛdwɛm] ‘I came out’, *przyszedłem* [pʃiʃɛdwɛm] ‘I came’, *wszedłem* [fʃɛdwɛm] ‘I came in’, *nadszedłem* [natʃɛdwɛm] ‘I arrived’, *podszedłem* [pɔtʃɛdwɛm] ‘I approached’, *zszedłem* [sʃɛdwɛm] ‘I descended’, *doszedłem* [dɔʃɛdwɛm] ‘I reached’, *odszedłem* [ɔtʃɛdwɛm] ‘I went away’, *obszedłem* [ɔpʃɛdwɛm] ‘I walked around’, *przeszedłem* [pʃɛʃɛdwɛm] ‘I crossed’, *zaszedłem* [zaʃɛdwɛm] ‘I dropped by’, *uszedłem* [uʃɛdwɛm] ‘I walked away; I escaped’, *rozszedłem (się)* [rɔʃɛdwɛm] ‘I divorced’, and the respective feminine forms: *poszłam* [pɔʃwam], *wyszłam* [viʃwam], *przyszłam* [pʃiʃwam], *weszłam* [veʃwam], *nadeszłam* [nadeʃwam], *podeszłam* [pɔdeʃwam], *zeszłam* [zeʃwam], *doszłam* [dɔʃwam], *odeszłam* [ɔdeʃwam], *obeszłam* [ɔbeʃwam], *przeszłam* [pʃeʃwam], *zaszłam* [zaʃwam], *uszłam* [uʃwam], *rozeszłam (się)* [rɔzeʃwam] <sup>71</sup>. In addition to the past tense paradigm, the suppletive stem occurs in the [ʃɛd] variant in the rarely used form of the past participle, cf. *szedłszy* ‘having gone’, *poszedłszy* ‘having gone (away)’, etc. I will exclude this form from further discussion.

Although the alternation is limited to a few verbs only, it constitutes a relatively strong pattern due to high text frequencies of the past tense forms representing both

<sup>71</sup> The *zero~e* alternation of the prefix vowel in some examples of masculine *versus* feminine forms is regular in this context and will not be discussed here.

allomorphs. All verbs in this “gang”, in Bybee’s (2001) terms, belong to the common, everyday lexicon and are probably much more frequent in the spoken language than reflected by the PWN Corpus figures. But even the latter, included for five sample verbs in Table XXIX, are significantly high for the 3<sup>rd</sup> person. The relatively low figures for the 1<sup>st</sup> and 2<sup>nd</sup> persons naturally result from the small percentage of oral texts included in the corpus material; otherwise, we would expect them to be much higher. Let us also observe that the masculine forms (listed as the first ones in groups joined by the colon) are about twice as frequent than the respective feminine forms, which in turn are slightly more frequent than the neuter ones. These correlations comply with cross-linguistic tendencies (cf. Greenberg 1966). The only exception is extremely high occurrence of the neuter form *doszło* (*do tego, że*), due to its grammaticalization into the meaning ‘it has got to the point (that), it happened (that)’.

Table XXIX. The PWN frequencies of 3<sup>rd</sup>, 1<sup>st</sup> and 2<sup>nd</sup> person singular masculine and feminine past tense forms.

szedł 1213: szła 628: szło 580, szedłem 191: szłam 100, szedłeś 9: szłaś 8  
 poszedł 1804: poszła 974: poszło 658, poszedłem 429: poszłam 271, poszedłeś 0:  
 poszaś 23  
 wyszedł 1983: wyszła 1281: wyszło 1100, wyszedłem 260: wyszłam 158,  
 wyszedłeś 16: wyszaś 28  
 wszedł 1756: wszła 948: wszło 363, wszedłem 201: wszłam 121, wszedłeś 13:  
 wszłaś 0  
 doszedł 935: doszła 413: doszło 5800, doszedłem 150: doszłam 78, doszedłeś 7:  
 doszaś 5

Table XXX summarizes the frequencies of all word-forms within the paradigms of the five sample verbs, according to their stem variant. Within the singulars, [ʃed] occurs slightly more often than [ʃ]. But if we include the plurals for comparison too, we can see that the total number of the occurrences of the [ʃ] allomorph is higher. In any case, however, the differences are not severe given the very high frequency of each form. Recall from chapter 5 that verbs of similar or even smaller frequencies tolerate paradigmatic allomorphy quite well. The particular case here apparently looks like a very stable pattern, too. The allomorphic template is supported by the existence of a small “gang” sharing the same stem. All members of the “gang” have high frequencies of WORD, as well as of word-forms of both allomorphs. In addition, the three inflectional forms of the minor [ʃed] variant share the feature of “masculine singular”, which enhances the salience of this allomorph. In sum, we do not expect the good “gang” of *iść* and its derivatives to be much susceptible to analogical leveling.

Table XXX. The total PWN frequencies of both stem allomorphs.

singular [ʃed]	singular [ʃ]	plural [ʃ]	total [ʃ]
szedł etc. 1413	szła etc. 1316	szli etc. 1427	2743
poszedł etc. 2242	poszła etc. 1565	poszli etc. 1610	3175
wyszedł etc. 2259	wyszła etc. 1467	wyszli etc. 1160	2627
wszedł etc. 1970	weszła etc. 1069	weszli etc. 1122	2191
doszedł etc. 1092	doszła etc. 6296	doszli etc. 886	7182

On the other hand, if there were a tendency for leveling, it seems that both allomorphs are possible Bases. For example, a natural scenario could involve leveling the feminine/neuter forms towards the more frequent masculine ones within the singular paradigm, i.e. hypothetical *?szedła*, *?poszedła* etc. This would eliminate the alternation in the singular and would maximize the opposition between singulars and plurals. This pattern, although quite stable, could later lead to the further expansion of the [ʃedw] allomorph in the plural, i.e. hypothetical *?szedli*, *?poszedli* etc., resulting in a complete removal of the alternation. An alternative direction of leveling assumes that the [ʃ] allomorph would replace [ʃed] in the 1<sup>st</sup> and 2<sup>nd</sup> persons, e.g. *szłem*, *szleś*, *poszłem*, *poszleś*. But there is a little problem with the 3<sup>rd</sup> person, which cannot be easily leveled. In the non-prefixed verb *szedł*, the leveling would make the verb vowelless, which is entirely impossible in Polish (cf. \*[ʃw]); in prefixed verbs, the leveling would produce a non-existing final coda, cf. the hypothetical *?poszł* [pɔʃw]. In spite of these difficulties, it is the latter direction of leveling which is expansive in the colloquial language, and there is no trace of the former strategy.<sup>72</sup>

### 6.3.2. The leveling of the [ʃ] stem in the colloquial language

It is my contention that if not for the fierce battle conducted by school teachers and language purists, the *poszłem/poszleś* type in the 1<sup>st</sup> and 2<sup>nd</sup> persons singular would have completely replaced the *poszedłem/poszedleś* type a long time ago. Children of all ages constantly use such forms, while being persistently corrected by their parents or teachers until they eventually undo it and learn the “correct” version. The use of the shorter, colloquial inflection varies greatly among grown-ups and is inversely proportional to their level of education. There are certainly persons who use only the colloquial forms, and most educated people attempt to use only the standard ones but occasional slips of the tongue may happen to them, too. The “incorrect” forms are not only spoken, but appear in writing as well, as can be easily concluded from a glance on the Polish Internet pages, flooded with them. The following figures come from Google search performed in March, 2007. The colloquial form *poszłem* appeared about 33.800 times, while its standard variant *poszedłem* had about 674.000 occurrences. Less frequent *szłem* occurred 19.000 times, while the standard form *szedłem* 214.000 times. Even if we assume that

<sup>72</sup> Apart from the joking situations, in which feminine forms as *szedłam* etc. can be used.

some of *poszłem/wyszłem* etc. cases are typographic mistakes and ironic uses, there will still be a large number of them left as just the normal way in which many people speak and write. And it seems that among these people there are such who otherwise use the standard variety of the language, free of other sub-standard features. Given that, we can conclude that the *poszłem* inflection type lodged itself into the standard language as a free variant. However, language purists and dictionary makers still refuse to recognize it as such, possibly because of the above-mentioned problem of the unlevelled form of the 3<sup>rd</sup> person singular (but as I will show momentarily, language users left without the help of linguists can cope with that, too). Nevertheless, the issue comes up from time to time in discussions by linguists as well as by common Polish speakers.

To sum up, there seems to be something very natural about the *poszłem* direction of leveling within this particular family of verbs in spite of the reasons mentioned earlier *against* such leveling. Recall that, firstly, words of such high frequencies do not have to level. In fact, this particular verbal stem is very irregular in all its inflection by having different suppletive stems in each of the three tenses, cf. *idę* ‘I go’, *szedłem/szłam* ‘I went’, *pójdę* ‘I will go’. Secondly, the other past tense allomorph provides in a sense a better Base for analogy, because it could be possibly leveled throughout the complete paradigm. Let us also observe that the *poszłem* direction of analogy does not seem to be motivated by a phenomenon known from early acquisition, when boys copy the speech of their mothers and use feminine verbal endings, e.g. *robiłam* ‘I did-fem.’ instead of the masculine *robiłem*. This process is easily undone when children become older and, to my knowledge, never continues to the adulthood.

I would like to suggest that the only meaningful explanation of the *poszłem* leveling relates to Zipf’s laws and the size of the standard versus colloquial forms. Within the standard paradigm, the masculine 1<sup>st</sup> and 2<sup>nd</sup> singulars are one syllable longer than the respective feminine and neuter forms<sup>73</sup>, although they are more frequent. The leveling solves the disproportion and provides a “size” appropriate to the frequent usage of these verbs. The choice of another allomorph from within the paradigm appears as a perfect strategy for an “irregular shortening caused by frequency”. Under this analysis, the satisfaction of the templatic requirement to differentiate masculine and feminine forms solely by the *e/a* vowel comes out as a nice by-product rather than the goal itself.

The colloquial extension of the stem allomorph [ʃ] onto the 1<sup>st</sup> and 2<sup>nd</sup> persons masculine singular leaves out only one form in the paradigm with the [ʃed] allomorph, that of the 3<sup>rd</sup> person masculine singular. As mentioned earlier, leveling of this form encounters phonotactic problems. Indeed, the speakers of *poszłem* normally use *poszedł* for the 3<sup>rd</sup> person. However, it seems that for some speakers, the prefixed 3<sup>rd</sup> person verbs as *poszł*, *wyszł* etc. have recently become a possibility. I, myself, have never heard such forms, nor anybody I asked about them has, but some cases are found on the Internet pages.<sup>74</sup> In (85) below, I present some examples coming from blogs written by

<sup>73</sup> With the exception of those verbs which contain an *e-zero* alternating prefix, cf. fn. 71, but these are less frequently used. More exactly, I would assume that the leveling first affects the most common verbs *poszłem*, *wyszłem* etc. and then, by analogy to them, less common *obeszłem*, *rozeszłem* etc., which are as long as the standard forms.

<sup>74</sup> I am grateful to Stanly Oomen who has told me (to my great disbelief) that such forms actually occur on Polish web pages, and who by discussing with me the *poszłem* type inspired me to write about it here. For

supposedly young people (judging by the contents of their blogs), which do not seem to be intentionally funny or ironic, but rather reflect their usual way of speaking. It should be mentioned that even though (85a) translates as the 1<sup>st</sup> person and (85b) as the 2<sup>nd</sup> person, both sentences contain the “participle” form *poszł*, because the clitics *-m* and *-ś* are attached to another word earlier in the sentence. In (85c), we can see that while the standard form is used for the less common verb *przyszedł*, the colloquial form *poszł* is used for the more common verb. As far as I can see, the standard 3<sup>rd</sup> person *szedł* is not threatened yet by any serious new proposal, although the Internet discussants are occasionally contemplating the versions as *szł* (rather unpronounceable in Polish) and *szoł*.

(85) The Internet *poszł* examples

- a/ też bym sobie poszł na piwko  
‘I would, too, have gone [to have] a beer’
- b/ jakbyś Wieśku poszł głosować to byś pewnie zauważył że [...]’  
‘if you, Wiesiek, had gone to vote, you would have noticed that [...]’
- c/ śnieg przyszedł i już poszł sobie  
‘Snow came and already went away’
- d/ i w jednej akcji Polituś wyszł z bramki to ja strzeliłem.  
‘and in one action [of the soccer game – IKS], Polituś came out of the goal, so I scored.’

---

those who would like to check the data themselves, I should note that numerous occurrences of sequences *poszł* etc. on the Internet are typos (intended *poszła/poszło*) and only some are intentional.

## CHAPTER 7

# UR-driven analogy

### 7.1. Introduction

According to Albright's (2002, 2005) explanation of analogy, leveling is oriented towards a form from within a paradigm which comprises the maximum of phonological information. The choice of such a Base prevents neutralization of phonological contrasts and is well motivated from the functional point of view. According to the author, the most informative member of a paradigm, acting as the Base, maximally "reveals" the underlying representation of the stem, so that other forms can be easily concluded from it. Such understanding of analogy is very appealing from the theoretical perspective, since Base selection can be done effectively, according to clear principles, and learning procedure can be adequately formulated, as well. Empirically, Albright's proposal certainly appears an appropriate analysis of some cases of leveling, to which I will return momentarily. However, for many others, e.g. all the Polish data discussed in chapters 2-4, it is inapplicable, since none of the stem variants in each of the three alternations would appear as more informative than the other. It seems that the cases described by Albright and their analysis involve a mechanism very different from analogy *per se*. An analogical process is activated first and foremost by the *need* to unify the stem due to a difficulty in maintaining the alternation, while the unification of the stem in cases described by Albright appears as a side-effect only. What is then its major trigger? Precisely what it seems, namely, the *need* to reveal the underlying representation, or, a mental image.

The process can be perhaps compared to hypercorrection with the only difference that the speaker aims not at an actually existing form of a socially higher dialect, but at a construed form, existing only in his/her mind. But in both cases, the speaker's strategy is oriented towards the hearer for whom a "better" form is projected. While the hypercorrected form is meant to impress the hearer, the analogical, UR-driven form usually serves the purpose of facilitating communication by making a word more transparent to the hearer. But in both cases, in order to produce a "better" form, the speaker must make an extra effort in suppressing the "other" form, known to him/her perfectly. On the other hand, analogy *per se* is triggered by the LAZY family of constraints, which minimize the speaker's effort, so that instead of retrieving a "correct" form, an easier substitute is produced. Here, unlike in hypercorrection and UR-driven analogy, the speaker is either unaware (unsure) of the existence of the other form or uses the two in variation.

Finally, the phenomenon is very similar to changes "affecting infrequent words first", as described by Phillips (1984), which do not involve a Base as an existing surface form. But the author convincingly links them to analogy, since, as she says: "both originate in the conceptual sphere of language, not in the articulatory parameters of the vocal tract" [Phillips 1984:336n]. In one of the cases discussed, glide deletion in South American English occurs more often in rare words, such as *nude*, *Tudor*, *tuber*, *tunic*, *neutron*, etc. than in frequent words as e.g. *new*, *knew*, *during* or *Tuesday*. The correlation between the frequency group and glideless pronunciation appears very strong (e.g. 43% for *new* and 74.4% for the rarest words). Given that rare words are mostly used in

writing, we can hypothesize that in this particular case, as well as some others (e.g. Polish loanwords discussed later), the orthographic convention of a written form helps to provide an idealized mental model at which the speaker will aim.

All examples discussed in the remaining parts of this chapter illustrate UR-driven analogy within inflectional paradigms. But let us observe that underlying representations are often more transparent in derived words, e.g. when a vowel-initial suffix “reveals” an underlying consonant neutralized word-finally in the morphological base. In such cases, we could very well expect UR-driven analogy to affect morphologically simpler forms.<sup>75</sup>

## 7.2. Historical (in Yiddish) and sporadic (in Polish) loss of final devoicing

One of Albright’s (2005) convincing examples of what I have called UR-driven analogy involves a very well known case of Yiddish, in which final devoicing was historically lost (cf. among others Anttila 1989/1972, Kiparsky 1968, Sadock 1973, Vennemann 1972).

The examples in (86) below, cited after Albright (2005:5), demonstrate that the process of final obstruent devoicing was attested in Yiddish at an earlier stage, but is no longer active in Modern Yiddish. In Earlier Yiddish, an underlying voiced obstruent contrasted with a voiceless one before a vowel-initial suffix, i.e. in most of declensional cases, while in the suffixless form of the singular nominative=accusative the contrast was neutralized due to final devoicing. In Modern Yiddish, the underlying voiced and voiceless consonants contrast in this environment, too. An analogy-based account of this change (unlike hypothetical final obstruent voicing) is supported by the data as (86b), in which an underlying voiceless consonant remained voiceless, as well as by the fact that voicing did not affect words as *vek* ‘away’, containing the same stem as the noun of (86a) but no plural with the voiced obstruent variant.

### (86) Suppression of final devoicing in Yiddish

		Earlier Yiddish		Modern Yiddish		gloss
		sg.	pl.	sg.	pl.	
(a)	nom., acc.	vek	veg(ə)	veg	vegən	‘way’
	gen.	vegəs	veg(ə)			
	dat.	veg(ə)	vegən			
(b)	nom., acc.	zak	zek(ə)	zak	zek	‘sack’
	gen.	zakəs	zek(ə)			
	dat.	zak	zekən			

<sup>75</sup> A possible example is reported by Joe Pater (available through Internet on *Phonoblog*), who observes that in his own dialect of English the *ŋ*-final adjectives as *long* are pronounced with the *ŋg*-sequence, presumably by analogy to derived forms as *lo[ŋg]er*. Since [ŋ] and [ŋg] contrast intervocally (*si[ŋ]er* versus *fi[ŋg]er*), we can claim that *lo[ŋg]* “reveals” the UR better.

Albright (2005), just as the previous researchers of the Yiddish case, assumes that the leveling of the voiced consonant in paradigms such as that of (86a) was towards the Base of the plural form, which is neither unmarked nor more frequent than the singular. In his account, this direction of the leveling is explicable since a voiced stem final obstruent is more informative, or, “underlying”.

In order to better understand the historical change in Yiddish, I would like to digress from this language and touch upon a similar phenomenon currently taking place in contemporary Polish. Contrary to Yiddish though, the suppression of final devoicing in Polish is sporadic, rather unpredictable and there is so far no indication that it is going to evolve into a more regular process or a historical change. Similarly to Yiddish, the phenomenon seems to be restricted to monosyllabic words only.

Final obstruent devoicing in Polish is found prepausally, cf. (87a), and before a word (of the same phonological phrase) starting with a vowel or sonorant (some dialects comprise nasals in this group and some do not), cf. (87b). Before an obstruent initial word, anticipatory assimilation in voice takes place, cf. (87c-d). Similarly to Early Yiddish, words with stems ending in an underlying voiced consonant preserve the voicing before a vowel initial suffix (as in (87e)) and a voiceless alternant appears (in the above-mentioned environments) only with zero suffix. In nouns, to which I will limit the present discussion, zero suffix appears in masculine singular nominative (=accusative of inanimates) and the plural genitive of feminines and neuters.

(87) Alternation of /sad/ ‘(fruit) garden’ (masc.)

a/ [sat]#	‘(fruit) garden-nom.’
b/ [sat ɔ]wocowy	‘fruit garden-nom.’
c/ [sat p]omarańczowy	‘orange garden-nom.’
d/ [sad v]iśniowy	‘cherry garden-nom.’
e/ [sadu]/[sadi] etc.	‘(fruit) garden-gen. sg./nom. pl.’

The facts of Polish described above are usually thought of as exceptionless. However, final devoicing is occasionally suppressed in monosyllabic words in the “devoicing” environments, such as (87a-b). The exact conditions of this suppression are rather hard to determine due to its relative rareness and unpredictability. Still, based on my observations of Polish speakers in actual language use situations, certain tendencies can be formulated as a hypothesis for further verification. I would like to stress, however, that the facts described here should not be confused with all kinds of pragmatic behaviors in which a word is overemphasized or has an exaggerated pronunciation due to impatient repeating, dictating, mocking etc., in which overdoing of voicing may take place, usually followed by a strong, vowel-like release. In the cases described here, no such special conditions are assumed, but rather normal intonation, with no particular emphasis or unusually strong release.

It seems that words most predisposed for the final devoicing suppression are those which are the least salient in terms of their frequency or phonological structure. Thus, for example, I have noticed that speakers of the television weather broadcast occasionally maintain the prepausal “phonemic” voicing in words, such as *wyż* [viʒ] ‘(atmospheric)

high' or *niż* [ɲiʒ] '(atmospheric) low', but I have not spotted similar voicing of coda in words, such as *śnieg* [ɕɲɛk] 'snow' or *mróz* [mrus] 'freezing temperatures'. While the two former words are of the specialized, rare vocabulary, the two latter are common in everyday usage. In addition, *śnieg* and *mróz* have complex onsets by which they have more salient structure than *wyż* or *niż*: final devoicing leaves out three fully recognizable segments of each of the former words, while it would leave out only two such segments in the case of the latter. Thus, even though all four words are fairly predictable to occur in the context of weather broadcast, the subtle difference among them with respect to their size and familiarity sometimes triggers the voicing strategy which provides an additional "phonemic" element of their structure to enhance the proper recognition of less salient words.

Foreign names of low familiarity are especially susceptible to final devoicing suppression. For example, I once observed on the news that the name of the Spanish river *Tag* was consequently pronounced as [tag] in the devoicing environment, even though it was mentioned several times and was then quite predictable from the earlier context. In this and similar cases, it can be hypothesized that the voiced, hypercorrected pronunciation is triggered by orthography which provides the "proper" UR and not by the voiced stem variant appearing in other declensional cases. Although foreign names like this one in principle can be inflected, in the real language situation they seldom are<sup>76</sup>, hence there are practically no cues as to the voiced stem final consonant apart from the orthography. Let us observe, that if this reasoning is accepted, we cannot talk about analogy to other forms of the paradigm, which supports my earlier argumentation against strictly analogy-based explanation of this phenomenon. Similarly to the case of the native vocabulary, suppression of the devoicing seems to be restricted to monosyllabic foreign words only.

Because the phenomenon of final devoicing suppression is sporadic and rare, the data are hard to obtain from a corpus. Therefore, I have run a small pilot test, whose intention was to slightly manipulate the speakers into using UR-revealing pronunciation, "empathic" towards the addressee. For the purpose of the experiment, a small „poem" of twenty lines was used. Each line of the poem consists of five syllables and ends with a monosyllabic rhyming word, which itself ends with an underlying sequence [-ud] (spelt as *-ud* or *-ód*). The consonant [d] occurs in a final devoicing environment (cf. the Annex for the full text and the details of the experiment). In some contexts, the final word of the line is fairly predictable, especially if it belongs to a fixed phrase as, e.g. *śłodka jak miód* 'sweet like honey', or *ósmymy świata cud* 'the eighth world miracle'. In other contexts, the final word comes as a little surprise, e.g. when it is an obsolete lexeme as *gród* 'town' or where the rare plural appears, e.g. *ideal wszech mód* 'an ideal of all the trends'. In still other contexts, the final word is quite unpredictable or even bizarre, e.g. *nizin, gór i wód* '[the people] of the lowlands, mountains and waters' or *majętny i z bud* 'the rich [people] and [those] of the shabby houses (lit. 'sheds' or 'doghouses')'. The rhyming words also differ as to their sizes, from the smallest, onsetless *ud* 'thighs-gen.' to as large as *w przód* 'ahead' or *w bród* 'plenty', having three consonants in the onset. My working hypothesis

---

<sup>76</sup> With foreign geographical names of low frequency one would tend to use fuller phrases, such as 'in the town of X', 'near the river of Y' (in which *X* and *Y* have invariant i.e. the nominative form) rather than 'in X', 'near Y' (in which *X* and *Y* occur with suffixes of other cases).

was that the voiced variant [-ud] will tend to appear in less predictable contexts to increase the salience of the word.

In the first version of the test, several native Polish speakers were individually asked to read the poem, while I was tape-recording them. But all of them maintained the final devoicing at the end of each line. Apparently, the stimulus of a word's unpredictability in the text was not sufficient enough: the speakers were reading the text for me, knowing that I was familiar with it. Therefore, there was no motivation for using an extra effort to suppress the final devoicing. In a modified version of the experiment, the participants were grouped into pairs of Readers and Addressees. None of them had seen the text of the poem before. Each Reader (R) was asked to read aloud the poem, while Addressee (A) was supposed to memorize as many of the rhyming words as possible and report them to me. I was tape-recording and writing down the words recalled by A, being thus a third participant – an addressee of A. The detailed results of this version of the experiment are included in Annex. They show that few Rs and As maintained final devoicing in all tested words. Most persons produced a fully or partially voiced [d] in 1-3 words in the less predictable contexts, but some Rs and As produced even more voiced codas than voiceless ones. The pilot test was conducted on a small group of 20 persons only and it is hard to determine exact factors which enhanced the hypercorrected, UR-driven pronunciation, apparently meant to be hearer-friendly. It seems that in addition to the purely linguistic parameter of the word's salience, the results depend more on such factors as type of personality, or a relationship (degree of familiarity) among participants rather than more easily measurable criteria, such as e.g. gender, age or birth place.

To conclude, the sporadic suppression of final devoicing in Polish is clearly hearer-oriented and aims at enhancing recognition of words whose small phonological structure and low frequency makes them dangerously indistinct. Whether a similar mechanism was at work at an early stage of the Yiddish change, from which it was further extended onto more common lexical items, remains at this point a hypothesis.

### 7.3. The development of Polish *deska* 'plank'

Within the range of Polish data discussed in the previous chapters, there are perhaps some for which it could be claimed that the leveling is UR-driven. A possible example is found in the development of the feminine noun *deska* 'plank', although I will show that alternative analyses are not unlikely, either.

Due to the presence of an earlier yer in the stem, the noun preserved the  $e\sim\emptyset$  alternation in its Old Polish form, which was further obscured by voicing assimilation and affrication, as illustrated in (88) below. After these changes had taken place, the vowelless stem allomorph [tsk] was found in all declensional cases except for the genitive plural whose stem was [desk]. But this minor alternant (in terms of the number of cases as well as frequency) was generalized within the whole paradigm eliminating the [tsk] allomorph. Subsequently, a new genitive plural, *desek* was formed, in compliance with the prevailing  $e\sim\emptyset$  alternating pattern found stem-finally.

(88) Stages of the development of *deska* ‘plank’

Early Polish	Old Polish	contemporary form+gloss
dъsk-a	dsk-a > tsk-a	deska ‘plank-nom.’
dъsk-ъ	desk	desek ‘planks-gen.’

The leveling of the allomorph *desk* could not be frequency-driven by simple majority criteria, even though for this particular noun, the genitive plural has exceptionally high frequency compared to rather rare average occurrence of this declensional case. What contributes to high frequency of the genitive plural is the expression *z desek* ‘from (of) planks’ used as a modifier in reference to floors, walls and other common objects made of wood shaped into planks. In the PWN Corpus, the genitive plural constitutes as much as 19,3% (253 tokens) of all 1314 occurrences of this noun (still, it is not the most frequent word-form, cf. *deski*, sg. gen. = pl. nom., with 410 occurrences i.e. 31,2% of total). In the sample of Old Polish data summarized in *Słownik* (1968-2004), the genitive plural occurs three times (7,9%) out of the total 38 occurrences.

We could perhaps account for these data by appealing to the fact that the *desk* allomorph fully reveals the underlying representation, while the allomorph *tsk* does not, and the leveling was thus towards the genitive plural form *desk*, in spite of its much lower frequency. However, there are additional facts which point to different possible scenarios of this historical change.

As Bańkowski (2000, vol. 1: 267) observes, a diminutive of this noun was used in Old Polish, with the regular phonological form *desfka* (in the nominative) or the simplified form *deska*. In the so-called Mazurian<sup>77</sup> dialects, which have dental fricatives and affricates instead of alveolar ones, this diminutive surfaced as *destska* or simplified *deska*. We can hypothesize with Bańkowski (2000) that the leveling was towards the stem *desk* revealed in the genitive plural *and* in diminutives. Or we could go even further and assume that *deska* was borrowed from the Mazurian dialects in the meaning of basic, non-diminutive form. Naturally, the existence of the same stem in “old” genitive plural could definitely encourage such borrowing and reinterpretation. Finally, we can assume that analogy and borrowing proceeded *together*, conspiring to the elimination of the uncomfortable alternation.

---

<sup>77</sup> These dialects spread much beyond the Mazurian region, hence in Polish they are called *mazurzańce*, which literary translates as “Mazuring” rather than “Mazurian”.

## CHAPTER 8

### Other issues

#### 8.1. The directionality of analogical mapping in derivational morphology

##### 8.1.1. The problem of derived Bases

In the OT literature on analogy, it is by and large assumed that if an O:O correspondence holds between a word and its morphological derivative, it is only the less complex word that may have a status of a Base, triggering the direction of the analogical mapping (cf. McCarthy 2005 and references therein). This conviction seems to be reminiscent of Kuryłowicz's second law of analogy, as well as of cyclic derivation in the pre-OT generative phonology. There certainly exists abundant empirical evidence pointing out to such "unidirectionality". Let us observe, however, that if derivational stages are eliminated as a possible strategy in the formal apparatus of OT and if empirical evidence to the contrary, even scarce, is found, unidirectionality loses its value as an unviolable principle.

Historically, more complex morphological words are formed on the basis of already existing, less complex words. The process of analogical mapping will thus naturally follow the path of making a new item (more complex, derived) resemble the old one (less complex, non-derived), which is already familiar and salient. However, analogy may come into play at the moment when a morphological derivative and its base are both "old" and well entrenched in language use. In this case, there should be no principled reason to favor the less complex morphological word as the only possible Base in analogical mapping, especially on the assumption that stem and pattern analogy use the same cognitive mechanism. Consequently, if pattern-analogy can extend a feature from a lexical item onto an unrelated lexical item and do so irrespectively from the words' morphological make-up, why should stem-analogy be limited by unidirectionality? I take it therefore for granted that there are no theoretical reasons that render a bi-directional mapping in derivational analogy *a priori* impossible.

Empirical evidence, however, demonstrates that derived Bases in analogy are exceptionally rare. I have already discussed a few possible cases in chapter 2. But generally speaking, it is hard to come up with unquestionable evidence, not evoking at least some reservation. In the remaining part of this chapter, I will argue that what looks like the unidirectionality principle actually results from a combination of other factors, related to language use. And that the joint outcome of these factors makes analogical mapping directed to the derived Base not entirely impossible but statistically very little probable. Thus, occasional cases are attested, but are not expected *en masse*.

One factor disfavoring derivatives' accessibility to the "Basehood" in a correspondence relation links to disproportionate frequencies of non-derived and derived words. From the semantic perspective, the former are typically more general than the latter (which have more specialized meanings) and as such, they have a wider scope of use. Assuming in addition that morphology is ubiquitously concatenative rather than non-concatenative (or truncative), non-derived words are normally phonologically less complex and shorter than derived ones. In accordance with Zipf's (1935) "frequency laws" discussed in chapter 6, shorter words statistically have higher token frequency than

longer words and vice versa. Let us recall from chapter 3 the frequency ranges of Polish nouns *vis-à-vis* their much less frequent “first” diminutives and exceptionally infrequent double diminutives with the reduplicated suffix. The directionality of analogy here goes hand in hand with the length of these nouns and their frequency of use. Examples of this kind are abundant, demonstrating that the usual analogical mapping from the morphologically less complex onto morphologically more complex often results from simple frequency criteria.

However, it may happen that, for whatever reason, the frequency of a derived word exceeds the frequency of the corresponding morphological base. But even in these statistically uncommon cases, there may be other obstacles for the derivative-to-base analogy, specifically, the templatic requirements of the morphological base category. I will illustrate the problem using a concrete example of English stress discussed in the following section. But for the purpose of that argumentation, firstly, it is important to recognize the fact that derivatives, even if very productive, constitute only a subset of all theoretically possible forms that could be derived from the existing morphological bases. In other words, a class of derivatives is always smaller than a class of non-derived (or less derived) bases. Let us again use the previous example of the diminutives discussed in chapter 3. In spite of the great productivity of the suffix *-ek*, out of the sample of 56 common nouns, fewer than the half (25) form acceptable diminutives with this suffix. Out of these, only few can form double diminutives with the reduplicated suffix. Similarly, the category of English verbs is larger in size than the category of derived nouns with the suffix *-ation*, and the latter is larger than the category of adjectives derived from them with the suffix *-al*, etc. The further, more elaborate the derivation, the smaller the lexical class that constitutes its output. These obvious quantitative relations have a number of consequences for analogy. For example, a small category will tend to avoid stem-allomorphy, as argued throughout this work, as well as resist pattern-allomorphy (cf. Mańczak’s differentiation law from chapter 1). Conversely, a large category will not welcome stem-analogy, if it were to destroy its well-established, recognizable templatic configuration.

### 8.1.2. Regular versus analogical stress in English derivatives

In this section, I will illustrate the problem of the hypothetical derived Base using the familiar example of English derivatives, lexically divided into two classes with respect to their stress/vocalism patterning. As Chomsky and Halle (1968) and many authors since observe, the middle secondary stress in words as e.g. *còndènsátion* is due to the fact that stress on this syllable appears in the base verb *condènze*. But this kind of stress “preservation” (in derivational terms) is not found in other derivatives, e.g. *cònversátion* versus *convèrse*. A non-derived OT analysis, as for example Pater’s (2000), attributes the difference between the two lexical classes to the difference in ranking between specific O:O correspondence constraints *vis-à-vis* general stress constraints, which reflects the intuition that the *còndènsátion*-type is analogical and the *cònversátion*-type is phonologically regular.

Various authors point out the connection between stress/vocalism preservation and frequency, as illustrated by the following quotations. According to Kager, variations are dependent upon “performance factors such as word frequency and the predictability

of words in discourse contexts” (1989:39). “Less frequent words tend to show their full vocalism presumably to aid the listener in accessing the correct lexical information” (Alcántara 1998:28, ft. 10). “[T]he establishment of words as relatively independent, and common, lexical items leads to their regularisation [...], on these grounds we would expect a correlation with the nominalisation facts, but would also expect other factors, such as frequency, to play a role” (Pater 2000:261). Hammond (1999) presents evidence that another well-known stress phenomenon of English, that of the “Rhythm Rule”, depends on frequency, triggering the main stress shift in common words, e.g. *àntique bóok*, but not in rare words, e.g. *àrcáne sórt*.

I have run a small frequency test using Pater’s (2000:238) data of 15 derivatives, selected by the author as to their morphological and phonological form so that the distinction between stress-preserving and non-preserving classes is purely lexical. I indicate the frequencies of both: the morphologically complex derived words and their bases, as they appear in the British National Corpus (the sample available through the Internet).

In most examples of non-preservation of stress, the derivatives have much higher frequencies than their respective bases, as in (89a), which complies with the fact of their phonological independence. The derivatives in (89b) have relatively low frequencies, but their verbal bases are extremely rare. The regular stress can be explained by hypothesizing that the bases are not salient enough to trigger analogy. However, in these particular cases it is also possible to argue that analogy actually does take place, but it is oriented towards much more frequent nominal Bases, *ségment* and *pígment*, which lack stress on the second syllable.

(89) Phonological stress in English derivatives

(a)	infor <sup>m</sup> ation (38327)	infor <sup>m</sup> (286)	
	con <sup>v</sup> ersation (5169)	con <sup>v</sup> erse (13)	
	con <sup>s</sup> ervation (3943)	con <sup>s</sup> erve (55)	
	con <sup>s</sup> ultation (2593)	con <sup>s</sup> ult (319)	
	tran <sup>s</sup> formation (1712)	tran <sup>s</sup> form (187)	
	con <sup>f</sup> irmation (1144)	con <sup>f</sup> irm (751)	
	tran <sup>s</sup> portation (553)	tran <sup>s</sup> port (21)	
(b)	ség <sup>m</sup> entation (247)	ség <sup>m</sup> ent (0)	cf. ségment (749)
	píg <sup>m</sup> entation (54)	píg <sup>m</sup> ent (0)	cf. pígment (223)

When a base has a high frequency and a derivative significantly lower frequency, as in (90a), analogy can enforce stress preservation. But the same effect is also observed when both the base and the derivative are relatively rare, as in the examples in (90b), even in the case of derivatives with comparable or slightly higher frequencies than those of their bases. This fact complies with the general idea that stem alternations tend to be eliminated in rare words. It should be noted, however, that examples of this kind are problematic, since on the one hand, the Base for analogy is not sufficiently salient, and on the other, the words are too infrequent to welcome allomorphy. This unresolved conflict causes cases of this kind to be more prone to variation (e.g. *con<sup>d</sup>em<sup>n</sup>ation* and *con<sup>d</sup>ensation* according to Burzio 1994:185).

(90) Analogical stress in English derivatives

(a)	àdvántágeous (372)	advántage (7220)
	àuthèntícítý (362)	àuthèntíc (824)
(b)	còndèmnátíon (443)	condémn (473)
	còndènsátíon (336)	condénsè (61)
	ìmpòrtátíon (164)	impórt (62)
	àugmèntátíon (53)	àugmènt (19)

The sample data in (89-90) above support the hypothesis that the lexical split of these English derivatives into the “regular” and “analogical” classes is highly correlated with the text frequency of particular words. Returning to the problem laid out in the previous section, let us consider again the data of (89a), which demonstrate that some derived nouns have considerably higher frequencies than their respective verbal bases. Especially striking are those cases in which verbs are situated at low frequencies, e.g. *convérse* (13 occurrences in the corpus) or *consérve* (55), while the derived nouns are extremely frequent, cf. *cònvèrsátíon* (5169) and *cònsèrvátíon* (3943). In such particular circumstances, we could expect the analogical mapping to apply from the much more frequent noun onto the infrequent verb. In phonological terms, that would mean the initial stress on the verb, either secondary, i.e. \**cònvérse*, or even the only one, i.e. \**cónverse*. To my knowledge, cases as such are not reported and seem to be very unlikely (but not entirely impossible).

Let us observe that the hypothetical analogical development of the verb stress to match that of the more frequent derived noun would be in conflict with the “templatic” iambic stress of the verbal category. The template “the initial syllable unstressed, the second syllable stressed” characterizes numerous disyllabic verbs, being a salient, recognizable feature of the category “verb”. In some cases, this template in addition carries the functional load of distinguishing between the category of verbs and nouns, cf. *repórt* versus *réport*, *prodúce* versus *próduce* etc. The hypothetical analogical stress would either completely ruin (as in \**cónverse*) or at least distort (as in \**cònvérse*) the templatic marker of these verbs.

In terms of Bybee’s model (Bybee 1985, 2001), the stress template constitutes the “connecting lines”, generalized as a pattern. The pattern has considerable “strength”, because of a relatively large number of verbs “connected” by it. The analogical stress assigned to a given verb would break its “connection” with all other verbs sharing this pattern for the purpose of establishing just one identity connection between this verb and its derivative.

In the OT framework, we can interpret these facts as a large number of O:O templatic correspondence constraints among pairs of various verbs (cf. Myers 1999). The analogical stress would induce violations of many O:O templatic constraints between one particular verb and others. This cost would have to be paid for the satisfaction of O:O stem identity.

The situation of English stress sketched above significantly differs from the Polish cases discussed in chapter 2, in which morphological derivatives presumably served as Bases for analogy. Recall that the *e~o/a* alternation in paradigms constitutes a “weak” pattern, due to the small size of the alternating class. This is precisely the reason why

stem-analogy is particularly welcomed within these paradigms. Neither has the alternation a functional character similar to the English verb/noun distinction, since each declensional case is marked by an overt suffix. The stem leveling satisfies multiple O:O identity constraints within the paradigm, while violating only a “weak” O:O templatic constraint. Normally, leveling is oriented towards a member (members) of the paradigm, because it constitutes the most natural Base for the semantic reasons. But a frequent derivative, semantically close, may serve the purpose equally well, as e.g. *biedny* ‘poor’ for *bieda* ‘poverty’ < \**biada*.

## 8.2. Phonological universal constraints or language-specific morphophonological templates?

The standard version of OT (cf. Prince and Smolensky 1993/2004, but also McCarthy and Prince 1993, 1994) underlines the universal character of phonological constraints and attributes differences among languages to differences in constraint ranking. The OT architecture has also instruments to restrict the given constraint to a particular morphological environment or to a lexical class, for example by means of alignment or lexical constraints. It seems that what mainstream OT linguists avoid or are not extremely happy with are cases of constraints which are language-specific and not particularly well phonetically motivated. But it also seems that a large number of constraints in various languages are just like that, half-way between completely idiosyncratic allomorphy and universal phonology. The Polish alternations discussed in chapters 2-4 provide such an example. Out of the three alternations, only *e~a/o* justifies a phonological explanation, but only if a constraint reverses the historical order of the change, as argued in chapter 2. The *o~u* alternation seems little transparent after the loss of vowel length for one reason, and in addition, with exceptional exemption of nasals from the conditioning environment of voiced segments. The *e~∅* alternation has the smallest amount of phonological motivation, especially since the “ghost” vowel is not the least sonorous one in the language inventory, which normally should be the case.

From the perspective of language use and acquisition, however, the little phonological conditioning of an alternation does not present any problem. The least phonologically transparent of the above-mentioned Polish alternations happens to be the most productive of all, unlike the phonologically best motivated one which is the most recessive. To a language speaker, whether an alternation is phonologically conditioned or not seems quite irrelevant, provided it is salient in language usage. Consequently, such a difference should be equally immaterial in a linguistic analysis. I mention this fact, because it seems that often a very abstract “phonological” analysis is enforced merely for the reason of being “superior” than an alternative analysis based on language-specific morphophonological requirements. The *e~∅* alternation provides a good example, with numerous earlier proposals of abstract underlying representation of the “ghost” vowel, as mentioned in chapter 4. If we follow the same track, we should have similar analyses of other alternations, i.e. with an abstract *a* vowel alternating with *e*, an abstract *o* vowel alternating with *e* (or, two abstract *e* vowels, alternating respectively with *a* and *o*), and several other abstract segments in order to account for the *o~u* alternation, as well as for other vowel alternations found in Polish and not discussed here. Apart from enormous complexity of such an analysis, it would still have to include some reference to a

“template”, as demonstrated in chapter 4. So why not use the template right from the beginning? My understanding of “template” is perhaps broader than currently used in the OT literature and covers all cases of a recognizable string or pattern which helps to identify a given category. It is simply a relation of synchronic proportional analogy, or, in terms of Burzio (2005), an “entailment”. Reinterpretation of the earlier phonology in morphophonological or “templatic” terms comes as a natural stage in language development in all languages with a relatively long history, i.e. perhaps with an exception of pidgins and creoles.

From the perspective of language change, when a rule loses its phonetic motivation and is no longer automatic, it may take a “weakening” or “strengthening” path. The former option characterizes a situation in which the alternation remains restricted to a limited set of lexical items and does not extend onto new vocabulary. In time, the scope of such a lexical rule further narrows down, since some words exhibiting the original alternation may come out of use and others may be leveled by analogy. Eventually, only a few “exceptions” are left as relics of an earlier productive rule before it completely dies out. The opposite situation is found in the case of a “strengthening” path. Even though no longer phonologically motivated, an alternation spreads onto new lexical items which share with the “old” ones the same morphophonological environment. In this work, I have referred to such an extension as pattern analogy. When such changes take place, the phonological conditioning of the original rule becomes more and more obscured, because analogy involves mapping between similar and not identical structures and because the regular phonological development occurs alongside (cf. Polish  $\bar{o} > u$ ). Eventually, obscure phonology becomes a clear template.

Automatic phonological rules often do not last too long. To use the Polish examples again, each of the historical processes: the Lechitic Vowel Shift, yer vocalization/deletion and vowel lengthening (raising) operated within the time span of no more than about 100-250 years. For many years since then, and presumably for many years to come, what is left of these alternations is analogy and a template. The latter does not mean reducing a linguistic analysis to listing language-specific alternations. On the contrary, throughout this work, I have tried to point out their meaningfulness by explaining their sources and motivation. The data discussed here provide evidence that the two opposite strategies of stem and pattern analogy directly relate to frequency. I have demonstrated that pattern analogy takes place when the size of an alternating category is relatively large and when it has members among high frequency words. Conversely, an alternation becomes unproductive when it is shared by few members of medium or low frequencies. There is also a strong correlation between maintaining an alternation, i.e. allomorphy and token frequency. Allomorphy positively correlates with high text frequency, while rare text frequency favors stem leveling. Anything else being equal, it is the more frequent allomorph (in terms of text frequency) which becomes the Base for leveling, and not a rarer one. Taking into account all such considerations provides an explanatory analysis, which consists not only in establishing the given hierarchy of constraints, but looks for its justification.

It is true that some alternations have a phonological rationale only from the historical perspective. But this should not be taken as evidence that a diachronic explanation must be better than a synchronic one. Language speakers, with the exception of those who happen to be linguists of course, have a very limited access to earlier forms

of their language (through conservative orthography, older literature etc.) and cannot rely on the logic of the old system. The constant, active work of analogy shows that language speakers seek new means to reinterpret old phonology in a new way, equally meaningful and based, among other things, on a salient template.

Finally, the importance of a template in a linguistic analysis does not undermine the role of universal phonology. The two simply coexist.

### 8.3. Concluding remarks

There is a big pressure in our time to produce “new” things, with an automatic implication that they are better than the old ones. While not necessarily sharing this opinion in all possible circumstances, I have written a relatively “old” thing here, since the main idea advocated in this work has been known in one way or another for over a century. As pointed out in chapter 1, the frequency-analogy connection was recognized at least as early as Kruszewski (1879) and since then has been supported by thorough exemplification and theoretical arguments in the research of Witold Mańczak, Joan Bybee and many others. So what new have I done then? I have attempted to show that in order to properly understand this connection, we must consider text and pattern frequency simultaneously, looking at the semantic side as well. I have also proposed how language use data can be incorporated into the formal apparatus of standard OT without putting the equal sign between “usage” and “grammar”. “Usage” does not automatically lead to “grammar” and should not be expected to, because in addition to being a host of physical and statistical laws, language is a social and cultural phenomenon.

An analogical change is slow and gradient – it affects single lexical items one by one and the process is extended in time. The cases of analogy discussed in chapter 2 provide a good illustration. Some of the nouns underwent analogical leveling before the 16<sup>th</sup> c., as in the case of e.g. *czas* ‘time’ or *cena* ‘price’, some others – as recently as in the 20<sup>th</sup> c., cf. *zwierciadło* ‘mirror’. We can expect with high probability that other lexical items will undergo leveling with respect to the *e~a/o* alternation in the future.

In the course of this work, I have sometimes indicated candidates for a possible analogical change to take place. For example, the locative=dative case of the word ‘sacrifice’ *ofierze*, or the locative=vocative of ‘flower’ *kwiecie*, discussed in chapter 2, can be expected to develop analogical forms *?ofiarze* and *?kwiacie*, respectively. Or, *torba* ‘bag’, the last alternating noun of the *-ba* ending class, discussed in chapter 4, may develop an analogical genitive plural *?torb* (present *toreb*) and diminutive *?torbka* (present *torebka*). Or, pattern analogy may affect the extremely frequent word *komputer* ‘computer’, as hypothesized in chapter 4, and shorten some of its forms by one syllable, so that the word’s size would better match its frequency, cf. the genitive *?komputra* (present *komputera*). On the other hand, we do not expect similar analogical changes to affect other lexemes, either because their stem alternation is entrenched by high text frequency, or because it is justified by a salient pattern of a strong, well represented category. With almost sure probability we do not expect that once leveling has affected a given word, the alternation will reappear in the same word. That is because the process of leveling takes place within a “weak” category (with a weak template) and an introduction of stem alternation involves a “strong” category with a salient template. And a template cannot be weak and strong at the same time.

But, as with all other kinds of language change, whether analogy will actually take place or not, even in the highly probable cases, can be estimated to the extent that no other factor intervenes. If computers suddenly disappear from our life due to the invention of an even better device called by a different name, pattern analogy will rather not have a chance to apply and the rare objects found in the museum will be designated by their long word-forms. If a popular song introduces a fancy phrase with the locative *kwiecie* ‘flower’, which will become a common saying and eventually will grammaticalize to become an everyday greeting, chances of leveling diminish greatly, etc. In short, even though we can predict which changes may happen, as Jerzy Kuryłowicz said “human factor decides whether and to what degree these possibilities become reality” [Kuryłowicz 1960:94, transl. I.K.S.].

An adequate model of grammar should be able to reflect the dynamic character of language. I have attempted to show in this work that language dynamics is strongly correlated with the lexicon, not in the narrow dictionary sense, but as real words occurring in language usage. Detailed examination of language use data contributes to our understanding of the broad-spectrum relation between lexicon and grammar. From the perspective of generative linguistics, phonology, whether modeled as rules or constraints, constitutes an active factor with the lexicon being passively affected by it. Generative phonology does not ask a question, or at least such questions are not at the heart of the mainstream research, why language A has the given rule set (constraint hierarchy) and language B a different one. In any case, I do not suppose that in the generative tradition, a possible response to such a question could be: “A and B have different phonologies because their lexicons differ”. But this, at least in the area of morphophonology, seems to be quite true, and the active role of the lexicon is implemented by analogy.

## ANNEX

### The “final devoicing suppression” experiment

For the purpose of this test a small „poem” was used<sup>78</sup>, consisting of four stanzas, each of five lines. Each line of the poem has five syllables and ends with a monosyllabic rhyming word, which itself ends with an underlying sequence [ud] (spelt as *ud* or *ód*). The consonant [d] occurs in a final devoicing environment due to the caesura (marked with a coma) and an additional precaution: each following line starts with a sonorant to avoid voicing assimilation in case a speaker does not make a pause (but all did).

Twenty persons participated in the experiment, including three children aged 11, 14 and 16. One person read the text of the poem (Reader), while another one (Addressee) was listening trying to memorize as many rhyming words as possible, which were then reported to me.

Jest słodka jak miód,  
Ma wdzięk, zgrabny chód,  
A piękno jej ud,  
Ideal wszech mód,  
Ósmy świata cud!

She is sweet like honey,  
She has charm, graceful walk,  
And the beauty of her thighs,  
[She is] an ideal of all the trends,  
The eighth world miracle.

Uwielbia ją lud,  
Nizin, gór i wód,  
Na wsi oraz gród,  
Majętny i z bud,  
Ludzki cały ród.

She is loved by the people,  
Of the lowlands, mountains and waters,  
In the countryside and [by] the town,  
[By] the rich and [those] of the shabby houses,  
[By] all human race (lit. family, clan).

Miłości tej trud,  
Jest jednak jak wrzód,  
Niemily jak głód,  
Albo trwały brud,  
I żrący jak sód.

The hardship of this love,  
Is, however, like an ulcer,  
Unpleasant like hunger,  
Or permanent dirt,  
And caustic like sodium.

Jej serce to chłód,  
O ile nie lód,  
O, ludzie, zmaż z bród,  
Łzy i ruszaj w przód,  
Innych panien w bród.

Her heart is coldness,  
If not ice,  
Oh, [you] folk, wipe off from [your] chins,  
The tears, and go ahead (i.e. move on),  
There is a plentiful supply of other maidens.

---

<sup>78</sup> I reluctantly admit to the authorship of this “poem”, but I also confess that it was somewhat influenced by my vague recollection of a similar thing which my two friends and I composed at the age of twelve or so.

The table below shows the results of the experiment. No instrumental measure has been used to determine the degree of voicing, since the contrast between the final [d] and [t] was very well audible. Only in a few cases, partially devoiced stop appeared, which I indicate in the table below as *dt*. Participants are grouped into pairs, e.g. (1-2), (3-4) etc., with an odd number indicating Reader (R) and the following even number - Addressee (A); F stands for “female” and M for “male”, and the following figure indicates the age of the person. The table does not include pairs 17-18 (RM 23, AF 30) and 19-20 (RF 48, AF 44), who produced only voiceless [t].

All participants live in Warsaw now and most of them come from Warsaw or its suburbs with the exception of: A6 (Katowice), R7 (Koszalin), R10 (Kielce), R 15, R 17 and A 20 (Lublin), A16 (Gdańsk) and R 19 (Lidzbark Warmiński). The relationships between participants are as follows:

mother and daughter: 1-2, 11-12

sister and brother: 13-14

colleagues: 3-4, 5-6, 7-8, 9-10, 15-16, 17-18, 19-20

Additional remarks:

In the case of the homonymic pair *lud~lód*, some As indicated which word they meant and some did not.

For A6 and A8 only voiced pronunciation is marked in the table, but these participants produced several words with [t] as well, which were impossible to identify, because of an incidental problem with the recording (voiced data were available from my notes).

R3 produced two partially devoiced codas at the beginning of the reading, but then she suddenly increased the speed and produced all fully devoiced codas.

R7 read *z bud* with the [t], but later on said it with the clearly voiced [d] (indicated in parentheses) helping the Addressee to recall it. Likewise R13 with respect to *ud*. I add these instances to the count of As, since they are citation forms.

After the recording, I usually (but not always) talked to the participants about the real purpose of the experiment. During those conversations, A10, who produced mostly voiced variants, told me that her boyfriend, a linguistics student, often pointed out to her that she tended to use hypercorrected pronunciation. R15, a student with some linguistic background (which I was unaware of), guessed (as the only person) that the purpose of the experiment was the pronunciation of the codas. She said that at the beginning of the recording, she realized she was producing voiced codas and tried to suppress it later (but cf. the results). This person read the text twice (because her partner A16 was somewhat confused about the task – actually some confusion remained after the second reading, too, cf. the “poor” results). During the second reading her attempt to suppress the voiced reading was more successful, but there was still one word (*glód*) pronounced with [d] (in the table only the first reading is marked).

	1 RF 46	2 AF 18	3 RF 24	4 AF 24	5 RF 24	6 AF 22	7 RM 23	8 AF 23	9 RF 22	10 AF 20	11 RF 42	12 AF 14	13 RF 11	14 AM 16	15 AF 24	16 RF 27	All Rs' d/dt	All As' d/dt
miód	t	t	t	t	t		t		t	d	t	t	dt	t	d	t	1dt	2d
chód	t	d	dt	t	t		t		t	-	t	t	t	t	t	t	1dt	1d
ud	t	t	t	t	t	d	t	d	t	d	t	t	t (d)	-	d	-	-	4d
mód	dt	-	dt	-	t		t		d	-	t	-	dt	-	t	-	1d 3dt	-
cud	t	t	t	t	t		t		t	d	t	t	t	-	t	-	-	1d
lud lód	t	t	t	t	t	d	t	t d	d	d	t	t	t	t	t	-	1d	3d
wód	t	d	t	-	t		t		t	-	t	-	t	-	d	-	-	2d
gród	t	t	t	-	t		t		d	-	t	t	t	t	t	-	1d	-
z bud	t	t	t	t	t		t (d)		d	-	t	-	t	-	t	-	1d	1d
ród	t	t	t	-	t		t		dt	-	t	-	t	-	t	-	1dt	-
trud	t	t	t	-	t		t		d	d	t	-	t	-	t	-	1d	1d
wrzód	t	t	t	-	t		t		t	-	t	t	t	-	t	-	-	-
glód	t	t	t	-	t		dt		t	-	t	-	t	-	d	-	1dt	1d
brud	t	t	t	t	t		t		t	t	t	-	t	t	t	-	-	-
sód	d	t	t	-	t		dt		t	-	t	-	t	t	t	-	1d 1dt	-
chłód	t	t	t	t	t		t	d	d	-	t	t	t	t	d	-	1d	2d
z bród	t	-	t	-	t		t	t	d	-	d	-	t	t	t	-	2d	-
w przód	t	-	t	-	t		t		t	d	d	-	t	t	t	-	1d	1d
w bród	t	t	t	-	t		t		t	-	d	-	t	t	t	-	1d	-

## REFERENCES

Almost all works listed as “manuscripts”, “Ph.D. dissertations” and “to appear” are available through the Internet on the authors’ web-pages or on Rutgers Optimality Archive (ROA) at <http://roa.rutgers.edu>.

- Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32. 239-289.
- Albright, Adam. 2002. *The identification of bases in morphological paradigms*. University of California, Los Angeles, Ph.D. dissertation.
- Albright, Adam. 2005. *Explaining universal tendencies and language particulars in analogical change*. MIT manuscript.
- Alcántara, Jonathan B. 1998. *The architecture of the English lexicon*. Cornell University Ph.D. dissertation. ROA-254.
- Anderson, Stephen R. 1988. “Morphological change”. In: Newmeyer (ed.). 324-362.
- Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam/Philadelphia: Benjamins [1<sup>st</sup> edition 1972].
- Anttila, Raimo. 2005. “Analogy: The warp and woof of cognition”. In: Brian D. Joseph and Richard D. Janda (eds.). 425-440.
- Anttila, Raimo and Warren A. Brewer. 1977. *Analogy: A Basic Bibliography*. Amsterdam: Benjamins.
- Bańkowski, Andrzej. 2000. *Etymologiczny słownik języka polskiego* [Etymological dictionary of Polish]. Vol. 1 and 2. Warszawa: Państwowe Wydawnictwo Naukowe.
- Batibo, Hermann M. and F. Rottland. 1992. “The minimality condition in Swahili word forms”. *Afrikanistische Arbeitspapiere* 29: 89-110.
- Baudouin de Courtenay, Jan Nicisław. 1904. *Szkice językoznawcze* [Linguistic essays]. Reprinted (1974) in: *Dziela wybrane*, vol. 1: 145-616. Warszawa: Państwowe Wydawnictwo Naukowe.
- Benua, Laura. 1995. “Identity effects in morphological truncation”. In: J. Beckman, L. Walsh-Dickey and S. Urbanczyk (eds.). *University of Massachusetts Occasional Papers in Linguistics: Papers in Optimality Theory*. 77-136. Amherst: GLSA.
- Benua, Laura. 1997. *Transderivational identity: Phonological relations between words*. University of Massachusetts Ph.D. dissertation. ROA-259.
- Brückner, Aleksander. 1974. *Słownik etymologiczny języka polskiego*. [Etymological dictionary of Polish]. Warszawa: Wiedza Powszechna [1<sup>st</sup> edition 1927].
- Burzio, Luigi. 1994. *Principles of English Stress*. Cambridge: Cambridge University Press.
- Burzio, Luigi. 2005. *Lexicon and grammar: unequal but inseparable*. Johns Hopkins University manuscript.
- Bybee, Joan. 1985. *Morphology: a study of the relation between meaning and form*. Amsterdam and Philadelphia: John Benjamins.
- Bybee, Joan. 1998. “The emergent lexicon”. *CLS* 34/2 (*The Panels*): 421-435.
- Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, Joan and Carol L. Moder. 1983. “Morphological classes as natural categories”. *Language* 59/2: 251-270.

- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago/London: University of Chicago.
- Caubet, Dominique. 1993. *L'Arabe Marocaine*. Vol. 1, *Phonologie et Morphosyntaxe*. Paris-Louvain: Éditions Peeters.
- Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Czaykowska-Higgins, Ewa. 1988. Investigation into Polish morphology and phonology. MIT Ph.D. dissertation.
- Danecki, Janusz. 1989. *Wstęp do dialektologii języka arabskiego* [Introduction to Arabic dialectology]. Warszawa: Uniwersytet Warszawski.
- Długosz-Kurczabowa, Krystyna and Stanisław Dubisz. 1998. *Gramatyka historyczna języka polskiego* [Historical grammar of Polish]. Warszawa: Uniwersytet Warszawski.
- Długosz-Kurczabowa, Krystyna and Stanisław Dubisz. 1999. *Gramatyka historyczna języka polskiego. Słowotwórstwo*. [Historical grammar of Polish. Word formation.]. Warszawa: Uniwersytet Warszawski.
- Furdal, A. 1964. *O przyczynach zmian głosowych w języku polskim* [On causes of sound changes in Polish]. Wrocław: Zakład Narodowy im. Ossolińskich.
- Greenberg, Joseph H. 1966. *Language Universals*. The Hague/Paris: Mouton.
- Grzegorzczkova, Renata, Roman Laskowski and Henryk Wróbel (eds.). 1984. *Gramatyka współczesnego języka polskiego. Morfologia* [Grammar of contemporary Polish. Morphology]. Warszawa: Państwowe Wydawnictwo Naukowe.
- Gussmann, Edmund. 1980. *Studies in Abstract Phonology*. Cambridge, Mass.: MIT Press.
- Hammond, Michael. 1999. „Lexical frequency and rhythm”. In: Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan and Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics*, vol. 1. 329-358.
- Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.
- Heine, Bernd. 1997. „Grammaticalization theory and its relevance to African linguistics”. In: Robert K. Herbert (ed.), *African Linguistics at the Crossroads. Papers from Kwaluseni. 1<sup>st</sup> World Congress of African Linguistics, Swaziland, 18-22 July, 1994*. 1-15.
- Henschel, Gert. 1996. “Zmiany fleksyjne a częstotliwość” [Inflectional changes and frequency]. *Studia Historycznojęzykowe II. Fleksja historyczna*. Kraków: Instytut Języka Polskiego PAN. 43-49.
- Hock, Hans H. 1986. *Principles of Historical Linguistics*. Berlin/NY/Amsterdam: Mouton de Gruyter.
- Hock, Hans H. 2005. “Analogical change”. In: Brian D. Joseph and Richard D. Janda (eds.). 441-460.
- Hooper, Joan [Bybee]. 1976. “Word frequency in lexical diffusion and the source of morphophonological change”. In: William M. Christie, Jr. *Current Progress in Historical Linguistics*. Amsterdam: North-Holland. 95-105.
- Hudson, Grover. 1980. “Automatic alternations in transformational phonology”. *Language* 56: 94-125.

- Humboldt, Wilhelm von. 1988. [1836-1839]. *On Language: The Diversity of Human Language Structure and Its Influence on the Mental Development of Mankind* [Translated from German by Peter Heath]. Cambridge: Cambridge University Press.
- Indeks a tergo do "Słownika Języka Polskiego" S. B. Lindego [A tergo index to S. B. Linde's "Dictionary of Polish"]. 1965. Doroszewski, Witold (ed.). Warszawa: Uniwersytet Warszawski.
- Jarosz, Gaja. 2005. Contextual correspondence and Polish vowel-zero alternations. Johns Hopkins University manuscript.
- Johnstone, T. M. *Eastern Arabian Dialect Studies*. 1967. London: Oxford University Press.
- Joseph, Brian D. and Richard D. Janda (eds.). 2005. *The Handbook of Historical Linguistics*. Malden: Blackwell.
- Kenstowicz, Michael. 1996. "Base identity and uniform exponence: Alternatives to cyclicity". In: Jacques Durand and Bernard Laks (ed.), *Current Trends in Phonology: Models and Methods*. Vol. 1. Salford: University of Salford. 363-394.
- Kenstowicz, Michael. 1997. "Uniform exponence: extension and exemplification". In: V. Miglio and B. Moren (eds.), *Selected papers from the Hopkins Optimality Workshop 1997, University of Maryland Working papers in Linguistics* 5: 139-54.
- Kenstowicz, Michael and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. San Diego: Academic Press.
- King, Robert D. 1969. *Historical Linguistics and Generative Grammar*. Englewood Cliffs, N.J.: Prentice-Hall.
- Kiparsky, Paul. 1968. "Linguistic universals and linguistic change". In: Emmon Bach and Robert T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston. 171-202. [Reprinted in Kiparsky 1988: 13-43]
- Kiparsky, Paul. 1972. "Explanation in phonology". In: Kiparsky 1988: 81-118.
- Kiparsky, Paul. 1978. "Analogical change as a problem for linguistic theory". In: Kiparsky 1988: 217-236.
- Kiparsky, Paul. 1982. *Explanation in Phonology*. Dordrecht: Foris.
- Kiparsky, Paul. 1988. "Phonological change". In: Frederick Newmeyer (ed.). 363-415.
- Kiparsky, Paul. 2000. "Opacity and cyclicity". *The Linguistic Review* 17:351-366.
- Kisseberth, Charles W. 1992. „Metrical structure in Zigula tonology". In: Derek F. Gowlett (ed.), *African Linguistic Contributions*. Pretoria: Via Afrika. 227-259.
- Kraska-Szlenk, Iwona. 1997. "Exceptions in phonological theory": Bernard Caron (ed.), *Proceedings of the 16<sup>th</sup> International Congress of Linguists*. Pergamon: Elsevier Science. CD ROM, Paper no 0173.
- Kraska-Szlenk, Iwona. 1999a. „Is analogy a threat to phonology?" Unpublished paper read at the 4<sup>th</sup> Holland Institute of Linguistics Phonology conference, January 28-30 1999, Leiden, Holland.
- Kraska-Szlenk, Iwona. 1999b. "Syllable structure constraints in exceptions". In: John R. Rennison and Klaus Kühnhammer (eds.), *Phonologica 1996: Syllables!?*, Proceedings of the Eighth International Phonology Meeting, Vienna 1996, Haga: Holland Academic Graphics, 113-131.
- Kraska-Szlenk, Iwona. 2003. *The phonology of stress in Polish. LINCOM Studies in Slavic Linguistics* 23. Muenchen: LINCOM Europa. [Based on the University of Illinois Ph.D. dissertation, Urbana-Champaign, 1995.]

- Kroesch, Samuel. 1926. "Analogy as a factor in semantic change". *Language* 2: 35-45.
- Kruszewski, Mikołaj. 1879. "Об 'аналогии' и 'народной этимологии'" (Volksetymologie)". *Русский филологический вестник*, vol. 2:109-120, Warszawa. [Reprinted in Polish in *Wybór pism*, Wrocław: Zakład Narodowy im. Ossolińskich, 1967:3-12]
- Krzyżanowski, Piotr. 1983. "Rola zmian analogicznych w procesach rozwojowych fleksji (na przykładzie wyrównań analogicznych tematów deklinacyjnych rzeczowników)" [The role of analogical changes in the development of inflection (exemplified by stem leveling in nominal declension)]. *Annales Universitatis Mariae Curie-Skłodowska. Sectio FF*, vol. I/8:109-122.
- Krzyżanowski, Piotr. 1992. *Temat fleksyjny w odmianie polskich rzeczowników* [The inflectional stem in the declension of Polish nouns]. Lublin: Uniwersytet Marii-Curie Skłodowskiej.
- Kuraszkiewicz, Władysław. 1972. *Gramatyka historyczna języka polskiego* [Historical grammar of Polish]. Warszawa: PZWS.
- Kuryłowicz, Jerzy. 1947. "La nature des procès dit analogiques". *Acta Linguistica* 5: 17-34.
- Kuryłowicz, Jerzy. 1960. *Esquisses linguistiques. (Prace Językoznawcze 9)*. Wrocław.
- Kuryłowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Heidelberg: Winter.
- Labov, William. 1972. "On the mechanism of linguistic change". In: Allan R. Keiler (ed.), *A Reader in Historical and Comparative Linguistics*. New York: Holt, Rinehart and Winston. 267-288. [Reprinted from: Georgetown University Monograph Series on Languages and Linguistics: Monograph No. 18, 1965:91-114]
- Labov, William. 1994. *Principles of Linguistic Change. Vol. 1 Internal Factors*. Oxford: Blackwell.
- Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago and London: The University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin/New York: Mouton de Gruyter.
- Laskowski, Roman. 1975. *Studia nad morfonologią współczesnego języka polskiego* [Studies on morphonology of contemporary Polish]. Wrocław: Zakład Narodowy im. Ossolińskich.
- Linde, Samuel B. 1860. *Słownik Języka Polskiego* [Dictionary of Polish]. Reprinted: 1994-1995, Warszawa: Gutenberg-Print, vol. 1-5.
- Malkiel, Yakov. 1967. "Each word has a history of its own". Reprinted (1983) in: *From Particular to General Linguistics. Selected Essays 1965-1978*. Amsterdam/Philadelphia: Benjamins. 217-226.
- Mańczak, Witold. 1958. „Tendences générales des changements analogiques”. *Lingua* 7: 298-325 and 387-420.
- Mańczak, Witold. 1965. *Polska fonetyka i morfologia historyczna* [Historical phonetics and morphology of Polish]. Warszawa: Państwowe Wydawnictwo Naukowe.
- Mańczak, Witold. 1966. "La nature du supplétivisme". *Linguistics* 28: 82-89.
- Mańczak, Witold. 1969. *Le développement phonétique des langues romanes et la fréquence*. Kraków.

- Mańczak, Witold. 1977. *Słowiańska fonetyka historyczna a frekwencja* [Slavic historical phonetics and frequency], *Prace językoznawcze* 55. Kraków: Uniwersytet Jagielloński.
- Mańczak, Witold. 1978. "Les lois du développement analogique". *Linguistics* 205: 53-60.
- Mańczak, Witold. 1980. "Laws of analogy". In: Jacek Fisiak (ed.), *Historical Morphology. Trends in Linguistics. Studies and Monographs* 17. The Hague: Mouton. 283-288.
- Mańczak, Witold. 1988. "O nieregularnym rozwoju fonetycznym spowodowanym frekwencją" [On irregular phonetic development caused by frequency]. *Biuletyn Polskiego Towarzystwa Językoznawczego* 41: 105-111.
- Mańczak, Witold. 1996. *Problemy językoznawstwa ogólnego* [Problems of general linguistics]. Wrocław: Zakład Narodowy im. Ossolińskich.
- McCarthy, John. 2005. "Optimal paradigms". In: L.J. Downing, T.A. Hall and R. Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press. 170--210.
- McCarthy, John and Alan Prince. 1990. "Prosodic morphology and templatic morphology". In: M. Eid and J. McCarthy (eds.) *Perspectives on Arabic Linguistics: Papers from the Second Symposium*. Amsterdam: Benjamins. 209-282.
- McCarthy, John and Alan Prince. 1993. "Generalized alignment". *Yearbook of Morphology*. 79-153.
- McCarthy, John and Alan Prince. 1994. "The emergence of the unmarked: Optimality in prosodic morphology. In: Mercè Gonzàles (ed.), *Proceedings of the North East Linguistics Society* 24. Amherst: Graduate Linguistic Student Association. 333-379.
- McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. In: Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics* 18: *Papers in Optimality Theory*. 249-384.
- Miachina, E. N. 1987. *Kamusi ya Kiswahili-Kirusi. Суахили-русский словарь*. Moscow: Russkij Jazik.
- Michalewski, Kazimierz. 1984. *Dystrybucja polskich rzeczownikowych formantów przyrostkowych* [The distribution of Polish derivational nominal suffixes]. *Folia Linguistica* 9. Łódź: Uniwersytet Łódzki.
- Myers, James. 1999. Lexical phonology and the lexicon. University of Manitoba manuscript. ROA-330.
- Newmeyer, Frederick J. 2003. "Grammar is grammar and usage is usage". *Language* 79/4: 682-707.
- Pater, Joe. 2000. "Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints". *Phonology* 17: 237-274.
- Pater, Joe. To appear. "The locus of exceptionality: Morpheme-specific phonology as constraint indexation". In : S. Parker (ed.), *Phonological Argumentation*. London: Equinox Publications. ROA-866.
- Phillips, Betty S. 1984. "Word frequency and the actuation of sound change". *Language* 60/2: 320-342.
- Piotrowski, Marek. 1992. "Polish yers and extrasyllabicity: An autosegmental account." In: Jacek Fisiak and Stanisław Puppel (eds.), *Phonological Investigations*. Amsterdam/Philadelphia: John Benjamins. 67-108.

- Piotrowski, Marek, Iggy Roca and Andy Spencer. 1992. "Polish yers and lexical syllabicity". *Linguistic Review* 92:27-67.
- Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden/Oxford: Blackwell. [Revision of 1993 Rutgers University Center for Cognitive Science technical report.]
- Retsö, Jan. 1983. *The Finite Passive Voice in Modern Arabic Dialects (Orientalia Gothoburgensia 7)*. Göteborg: University of Göteborg.
- Rospond, Stanisław. 2003. *Gramatyka historyczna języka polskiego* [Historical grammar of Polish]. Warszawa: Państwowe Wydawnictwo Naukowe [4<sup>th</sup> edition, 1<sup>st</sup> edition 1969].
- Rowicka, Grażyna. 1999. *On Ghost Vowels. A Strict CV Approach*. The Hague: Holland Academic Graphics.
- Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology. The Structure of Polish*. Dordrecht: Foris.
- Rubach, Jerzy. 1986. "Abstract vowels in three dimensional phonology: the yers". *Linguistic Review* 5: 247-280.
- Rubach, Jerzy and Geert E. Booij. 1990. "Syllable structure assignment in Polish". *Phonology* 7: 121-158.
- Rudzka-Ostyn, Brygida. 2000. *Z rozważań nad kategorią przypadku* [Deliberations on the category of case]. Kraków: Universitas.
- Russell, Kevin. 1995. Morphemes and candidates in Optimality Theory. University of Manitoba manuscript. ROA-44.
- Sadock, Jerrold M. 1973. "Word-final devoicing in the development of Yiddish". In: Braj B. Kachru, Robert B. Lees, Yakov Malkiel, Angelina Pietrangeli and Sol Saporta (eds.), *Issues in Linguistics. Papers in Honor of Henry and Renée Kahane*. Urbana/Chicago/London: University of Illinois Press. 790-797.
- Saeed, Jan. 1987. *Somali Reference Grammar*. Wheaton: Dunwoody Press.
- Schuchardt, Hugo. 1972 (1885). "On sound laws: Against the Neogrammarians" [transl. from German by Theo Vennemann and Terence H. Wilbur]. In: Theo Vennemann and Terence H. Wilbur (eds.). 39-72.
- Skousen, Royal. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer Academic Publishers.
- Sławski, Franciszek. 1952-1974. *Słownik etymologiczny języka polskiego*, [Etymological dictionary of Polish]. Vol. 1-5. Kraków: Towarzystwo Miłośników Języka Polskiego.
- Słownik frekwencyjny polszczyzny współczesnej* [Frequency dictionary of contemporary Polish]. 1990. Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor, Krzysztof Szafran, Jerzy Woronczak (authors) and Zygmunt Saloni (ed.). Kraków: PAN, Instytut Języka Polskiego.
- Słownik polszczyzny XVI wieku* [The dictionary of the 16<sup>th</sup> century Polish]. 1968-2004 (vol. 1-32). Instytut Badań Literackich PAN (ed.). Wrocław: Zakład Narodowy im. Ossolińskich.
- Sobelman, Harvey and Richard S. Harrell (eds.). 1963. *A Dictionary of Moroccan Arabic: English-Arabic*. Washington D.C.: Georgetown University Press.
- Spencer, Andrew. 1986. "A non-linear analysis of vowel-zero alternations in Polish". *Journal of Linguistics* 22: 249-280.
- Steriade, Donca. 2000. Lexical conservatism and the notion *base affiliation*. MIT

manuscript.

- Szpyra, Jolanta. 1992. "Ghost segments in nonlinear phonology: Polish yers". In: *Language* **68/2**:277-312.
- Tokarski, Jan. 2001. *Fleksja polska* [Polish inflection]. Warszawa: Państwowe Wydawnictwo Naukowe. [3<sup>rd</sup> edition]
- Traugott, Elizabeth C. and Bernd Heine (eds.). 1991. *Approaches to Grammaticalization*. Amsterdam: Benjamins, 2 vol.
- Ułaszyn, Henryk. 1956. *Ze studiów nad grupami spółgłoskowymi w języku polskim* [Studies on consonantal clusters in Polish], *Prace Językoznawcze* **8**. Wrocław: Zakład Narodowy im. Ossolińskich.
- Vennemann, Theo. 1972. "Phonetic analogy and conceptual analogy". In: Vennemann, Theo and Terence H. Wilbur (eds.). 181-204.
- Vennemann, Theo and Terence H. Wilbur (eds.). 1972. *Schuchardt, the Neogrammarians and the Transformational Theory of Phonological Change*. Bad: Athenäum.
- Wang, William S.-Y. (ed.). 1977. *The Lexicon in Phonological Change*. The Hague: Mouton.
- Zipf, George. 1965 [1935]. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Cambridge, Mass.: MIT Press.