

**The principle of least effort within the hierarchy of
linguistic preferences: external evidence from
English**

Małgorzata Kul

Praca doktorska
napisana pod kierunkiem
prof. dr hab. Katarzyny Dziubalskiej-
Kołączyk
w Instytucie Filologii Angielskiej
Uniwersytetu im. Adama Mickiewicza

Poznań 2007

Acknowledgements

The thesis has been made possible thanks to the School of English, Adam Mickiewicz University, which granted me the PhD scholarship in 2002. But the thesis would have been impossible without its spiritus movens, the head of the School and my supervisor, Prof. Katarzyna Dziubalska-Kołaczyk, whose continuous intellectual support guided me throughout the process of the thesis completion. In fact, Prof. Katarzyna Dziubalska-Kołaczyk has turned out to be somebody more than just a supervisor. In the times when I really needed it, she acted as a mother to me out of the goodness of her heart and for her invaluable help as well as for the spontaneous and wonderful help of all the School of English staff, I remain forever indebted.

Table of contents

Table of contents.....	3
List of tables.....	7
List of figures.....	8
List of abbreviations	10
Introduction.....	11
Chapter One: The principle of least effort: origins and meaning	14
1.1. The aim of the chapter	14
1.2. Review of terminology	15
1.3. The idea behind the principle.....	15
1.4. Operation of the principle	17
1.5. Typology of effort: the speaker's and the listener's	20
1.6. Review of literature	21
1.7. Motivation of the principle	27
Chapter Two: Hierarchy and preference in Natural Linguistics: epistemology of the concepts	35
2.1. The aim of the chapter	35
2.2. Natural Linguistics.....	35
2.3. The concept of a hierarchy.....	38
2.4. The concept of a preference	41
2.5. Preference-based theories in linguistics.....	44
2.5.1. Stampe (1973).....	44
2.5.2. Dressler (1997)	45
2.5.3. Vennemann (1983, 1988)	48
2.5.4. Dziubalska-Kořaczyk (2001)	51

Chapter Three: Articulatory effort.....	53
3.1. The aim of the chapter	53
3.2. The notions of easy and difficult sounds	53
3.3.1. Whitney (1878 [1971]).....	59
3.3.2. Baudouin de Courtenay (1877 [1972])	60
3.3.3. Zipf (1935 [1969])	61
3.3.4. Lindblom (1990)	63
3.3.5. Boersma (1998).....	65
3.3.6. Maddieson (p.c.)	66
3.4. Articulatory gestures	68
Chapter Four: Effort management: the experiments.....	72
4.1. The aim of the chapter	72
4.2. The principle of least effort as effort management.....	73
4.2.1. Explanation of effort management	73
4.2.2. The predictions of the experiments.....	74
4.2.3. Methodology of the experiments	75
4.2.4. The measure of articulatory effort	77
4.2.5. Discussion of effort parameters	78
4.2.6. The status of the parameters	80
4.3. The statistical aspect of the experiments	80
4.3.1. The method of establishing the sample size	80
4.3.2. Confidence interval.....	84
4.4. The empirical evidence	85
4.4.1. The results of the first experiment	86
4.4.2. The results of the second experiment	97
4.4.3. The attention hypothesis	103
4.4.4. The results of the third experiment.....	104
4.5. Discussion of the results	108
4.6. Effort management in text messages	110
4.6.1. Omission of punctuation.....	111
4.6.2. The use of small letter instead of capitals in proper names.....	111
4.6.3. The lack of spaces between words and the use of capitals to mark word beginnings.....	111
4.6.4. Emoticons	112

4.6.5. Onomatopoeic expressions	112
4.6.6. Replacement of the Polish characters with the ASCII ones	112
4.6.7. Borrowings.....	113
4.6.8. Number homophones.....	114
4.6.9. Letter homophones	114
4.6.11. Phonetic respellings	115
4.6.12. Clippings.....	116
4.6.13. Initializations	116
Chapter Five: Phonological processes in Natural Phonology.....	118
5.1. The aim of the chapter	118
5.2. The general nature and motivation of phonological processes.....	119
5.3. Phonological processes in language acquisition.....	122
5.4. The organization of phonological processes.....	125
5.5. The semiotic principles and parameters in phonological processes	128
5.6. Luschützky (1997)	138
Chapter Six: Phonological processes and the principle of least effort	142
6.1. The aim of the chapter	142
6.2. Typology.....	143
6.3. The current approaches to lenition/fortition	145
6.3.1. The traditional approach	145
6.3.2. The NP approach	147
6.3.3. The OT approach	149
6.3.4. The evaluation of the current approaches to process typology.....	150
6.4. The definition and criteria of lenition and fortition proposed in the thesis	154
6.4.1. Definition of lenition and fortition proposed in the thesis.....	154
6.5. Revision of process typology under lenition criteria	159
6.5.1. Segment deletion/ cluster reduction/simplification	161
6.5.2. Degemination.....	161
6.5.3. Monophthongization.....	162
6.5.4. Smoothing.....	162
6.5.5. Centralization.....	162
6.5.6. Epenthesis	163
6.5.7. Vowel insertion.....	164
6.5.8. Hiatus avoidance via linking or intrusive /r/.....	164

6.5.9. Diphthongization	164
6.5.10. Lengthening	165
6.5.11. Shortening	165
6.5.12. Assimilation of stops and nasals	165
6.5.13. Palatalization.....	166
6.5.14. Voicing.....	166
6.5.15. Devoicing of final obstruents.....	166
6.5.16. Debuccalization	167
6.5.17. Flapping	167
6.5.18. Stopping	168
6.5.19. Aspiration.....	168
6.5.20. Spirantization	168
6.5.21. Gliding	169
6.6. Revision of process typology under effort management	170
6.7. Teleological perspective	176
Chapter Seven: Conclusions	181
References.....	184

List of tables

Table 1. Modified coefficient of variation (subject 1).....	82
Table 2. Modified coefficient of variation (subject 2).....	83
Table 3. Quantiles $t(p, v)$ of p order of Student's distribution with v freedom degree....	84-85
Table 4. Vowel height.....	88
Table 5. Vowel duration in subsequent tokens of <i>key</i> [msec] (means of 12 speakers)...	99
Table 6. Duration of the vowel /i/ in subsequent tokens of <i>key</i> [msec] (speaker 3).....	100
Table 7. Repetition words from MICASE (office hours).....	106
Table 8. Processes vs. rules (after Katarzyna Dziubalska-Kołaczyk, p.c.).....	121
Table 9. Process types in prosodic and segmental domains (after Luschützky 1997: 68)..	139
Table 10. The current typology of phonological processes.....	160
Table 11. The revised typology of phonological processes under three lenition criteria...	169
Table 12. The revised typology of phonological processes under effort management.	173

List of figures

Figure 1. An illustration of a hierarchy - the healthy eating pyramid (after The Harvard School of Public Health 2004).....	39
Figure 2. An illustration of the phonological hierarchy (after SIL International 1999).....	41
Figure 3. The explanatory system of Natural Linguistics (after Dziubalska-Kołaczyk 2002b:104).....	45
Figure 4. Modified coefficient of variation (speaker 1).....	82
Figure 5. Modified coefficient of variation (speaker 2).....	83
Figure 6. Vowel duration.....	86
Figure 7. Confidence interval for vowel duration.....	87
Figure 8. F1 frequencies.....	88
Figure 9. Vowel height.....	88
Figure 10. Confidence interval for Figure 8.....	89
Figure 11. F2 frequencies.....	89
Figure 12. Confidence interval for Figure 11.....	90
Figure 13. F3 frequencies.....	91
Figure 14. Confidence interval for Figure 13.....	92
Figure 15. Closure duration.....	93
Figure 16. Confidence interval for Figure 15.....	94
Figure 17. Ratio of rise to total duration for the following vowel.....	95
Figure 18. Ratio of fall to total duration for the preceding vowel.....	95
Figure 19. Vowel duration in tokens used in the same context [msec].....	96
Figure 20. Vowel duration.....	97
Figure 21. Vowel duration.....	99
Figure 22. Duration of the /i/ vowel in subsequent tokens of <i>key</i> [msec].....	100

Figure 23. Confidence interval for Figure 22.	101
Figure 24: Normalized confidence interval for Figure 22.	101
Figure 25. Confidence interval for Figure 21.	102
Figure 26. Normalized confidence interval for Figure 21.	102
Figure 27. Duration of the /i/ vowel in subsequent tokens of key [msec] for the attention hypothesis.	104
Figure 28. Duration of 30 words.....	106
Figure 29. Confidence interval for Figure 28.	107
Figure 30. The scale of speakers' preferences with reference to articulatory effort.	109
Figure 31. The hierarchy of speakers' preferences with reference to articulatory effort.	109
Figure 32. Proportion of deleted vowels to consonants.....	115
Figure 33. A quintuple: (I) universals (II) type (III) language-specific competence (IV) sociolinguistic norms (V) performance (after Dressler 1985b: 292, Coseriu 1968)	121
Figure 34. A visualization of the semiotic principle of figure-and-ground (website Figure/ground at http://www.psychologie.tu-dresden.de/).	130
Figure 35. Labelled figure-and-ground effect (website Figure/ground at http://www.psychologie.tu-dresden.de/).	130
Figure 36. A picture representing the semiotic principle of figure-and-ground (website Figure/ground at http://www.psychologie.tu-dresden.de/).	130
Figure 37. A vase representing the semiotic principle of figure-and-ground (website Figure/ground at http://www.psychologie.tu-dresden.de/).	131
Figure 38. The concept of male and female in terms of the symbol - index - icon scale.	134
Figure 39. An illustration of sonority scale (after SIL International 1999).	150
Figure 40. A hierarchy of lenition criteria.	157

List of abbreviations

- C - consonant
- CD – constriction degree
- CL – constriction’s location
- EMMA – Electromagnetic Midsagittal Articulography
- EPG – electropalatography
- H&H – the Hyper- and Hypospeech theory
- MNP – Modern Natural Phonology
- MRI – Magnetic Resonance Imaging
- NL - Natural Linguistics
- NP - Natural Phonology
- OSDP – Optimal Sonority Distance Principle
- OT – Optimality Theory
- SMS – Short Message Service
- SPA – Stanford Phonology Archiving
- UP – Uniqueness Point
- UPSID – UCLA Phonological Segment Inventory Database
- V – vowel

Introduction

The thesis is an investigation of the principle of least effort (Zipf 1949 [1972]). The principle is simple (all effort should be least) and universal (it governs the totality of human behavior). Since the principle is also functional, the thesis adopts a functional theory of language as its theoretical framework, i.e. Natural Linguistics. The explanatory system of Natural Linguistics posits that higher principles govern preferences, which, in turn, manifest themselves as concrete, specific processes in a given language.

Therefore, the thesis' aim is to investigate the principle of least effort on the basis of external evidence from English. The investigation falls into the three following strands: the investigation of the principle itself, the investigation of its application in articulatory effort and the investigation of its application in phonological processes. The structure of the thesis reflects the division of its broad aims. The first part of the thesis presents its theoretical background (Chapter One and Chapter Two), the second part of the thesis deals with application of least effort in articulatory effort (Chapter Three and Chapter Four), whereas the third part discusses the principle of least effort in phonological processes (Chapter Five and Chapter Six).

Chapter One serves as an introduction, examining various aspects of the principle of least effort such as its history, literature, operation and motivation. It overviews various names which denote *least effort*, explains the origins of the principle and reviews the literature devoted to the principle of least effort in a chronological order. The chapter also discusses the nature and operation of the principle, providing numerous examples of the principle at work. It emphasizes the universal character of the principle from the linguistic field (low-level phonetic processes and language universals) and the non-linguistic ones (physics, biology, psychology and cognitive sciences), proving that the principle governs human behavior and choices.

Chapter Two provides the theoretical background of the thesis in terms of its theoretical framework and discusses the terms used in the thesis' title, i.e. *hierarchy* and

preference. It justifies the selection of Natural Linguistics as the thesis' theoretical framework by outlining its major assumptions and demonstrating its explanatory power. As far as the concepts of hierarchy and preference are concerned, the chapter provides their definitions and reviews their various understandings via decision theories and linguistic preference-based theories.

Since the thesis investigates the principle of least effort in language and speech, Chapter Three considers the articulatory aspect of effort. It reviews the notion of *easy* and *difficult sounds* and discusses the concept of *articulatory effort*, overviewing its literature as well as various understandings in a chronological fashion. The chapter also presents the concept of *articulatory gestures* within the framework of Articulatory Phonology.

The thesis' aim is to investigate the principle of least effort on the basis of external evidence, therefore Chapters Four and Six provide evidence in terms of three experiments, text message studies (Chapter Four) and phonological processes in English (Chapter Six).

Chapter Four contains evidence for the principle of least effort in articulation on the basis of experiments. It describes the experiments in terms of their predictions and methodology. In particular, it discusses the adopted measure of effort established by means of the effort parameters as well as their status. The statistical methods of the experiments are also clarified. The chapter reports on the results of the experiments, presenting them in a graphical way and discusses their relation to the tested predictions. Chapter Four establishes a hierarchy of speakers' preferences with reference to articulatory effort (Figures 30, 31).

The thesis investigates the principle of least effort in phonological processes, thus Chapter Five is devoted to the discussion of phonological processes in Natural Phonology. The chapter explains the general nature and motivation of processes as well as the development of processes in child language. It also discusses the organization of processes in terms of their typology as well as the order in which processes apply. The chapter characterizes the semantic properties of processes and overviews Luschützky's (1997) contribution to NP with respect to processes in terms of their typology and incorporation of articulatory gestures in the concept of a process.

Chapter Six investigates phonological processes. In particular, it identifies the issues of lenition/fortition definition and process typology by presenting the current approaches to process definitions and their typology. Since the chapter concludes that

no coherent definition of lenition/fortition exists, it develops alternative lenition/fortition definitions. The chapter also revises the typology of phonological processes under effort management, which is an extended version of the principle of least effort.

Chapter Seven concludes the thesis with a list of the concepts discussed in the thesis, enumerates the proposals made by the thesis in discussing the concepts and presents some questions for future research which have emerged in the course of investigation. The chapter also specifies the extent to which the investigation of the principle of least effort is a meaningful contribution to phonology.

Chapter One

The principle of least effort: origins and meaning

1.1. The aim of the chapter

The aim of the chapter is to describe the nature of the principle of least effort by presenting its several aspects. First, the chapter addresses the issue of terminology. Section 1.2. overviews various names which denote *least effort* in the literature. Next, the chapter defines selected concepts that underlie the idea of least effort. Section 1.3. presents Zipf's (1949 [1972]) discussion of the principle. Then, the chapter reports on the operation of the principle and notes that it governs word frequency as well as many other variables. Section 1.4. demonstrates the principle at work. Then, the chapter outlines the typology within the principle. Section 1.5. introduces the division into the speaker's and the listener's economy. Next, the chapter reviews the literature and reports on the research. Section 1.6. presents predecessors and followers of Zipf (1949 [1972]) in a chronological order. Finally, the chapter discusses the motivation of the principle. Section 1.7. gathers evidence concerning the principle from non-linguistic (physics, biology, psychology and cognitive sciences) and linguistic fields (low-level phonetic processes and language universals).

1.2. Review of terminology

A far as terminology is concerned, there exists a whole array of terms in the literature, used to refer to the same phenomenon or in an approximate sense. This array appears to result from the fact that different scholars from various fields of science have developed and employed their own terminology rather than from different interpretations of the least effort. This section attempts to compile a list of the names which refer to *least effort*. The most conventional and general term, i.e. *least effort*, seems to be the most adequate, or, the least vague and narrow. Other names of the principle include *least action* (Maupertuis 1750), *law of economy* (Whitney 1878 [1971]), *language economy* (Martinet 1960), *economy of effort* (Whitney 1878 [1971]), *tendency to ease* (Whitney 1878 [1971]) and *tendency towards convenience* (de Courtenay 1974). Boersma (1998) used the name *minimal effort*, whereas Bussmann (1996) called it *law of least effort*. *Maxima* and *Minima* are the names employed by Gengerelli (1930).

1.3. The idea behind the principle

Zipf (1949 [1972]) was the first scholar to explicitly formulate and formalize the principle of least effort.¹ He earned his reputation, however, as the author of the so-called Zipf's Law. It states that if one lists all the words of a language by how often they are used, the second most frequent word is about half as frequent as the most frequent one, the third most frequent is about a third as frequent as the most frequent one, the fourth is a fourth as frequent and so on.² The law was formulated on the basis of the findings of Yule (1944), Pareto (1897) and Estoup (1916) and modified by Mandelbrot (1965) who added two new adjustable constants and substituted fractions with their squares or cubes. Generally speaking, the law scales probability variables connected

¹ George Kingsley Zipf (1902-1950) graduated from Harvard in 1924 and has occupied the Professor Chair there for twenty years. He commenced his career as a philologist. However, with the course of time he described himself as a 'statistical human ecologist'. He was the first scholar who explicitly formulated the principle of least effort in his book *Human Behavior and the Principle of Least Effort*. This book was published in 1949 and proved a success with the readers.

² The frequency top ten in English comprises the following words: 1) the 2) of 3) to 4) a 5) and 6) in 7) that 8) for 9) was 10) with.

with size. Consequently, Zipf (1949 [1972]) applied the law to the following fields: geography, intranational and international conflict resolution, the distribution of economic power and social status, prestige symbols and cultural vogues. The task undertaken by him was enormous and he admitted that “this book (...) has been nearly six years in writing” (Zipf 1949 [1972]: x). Nowadays, the range of Zipf’s Law applications includes, among others, finance and business, web access, statistics and infometrics.

Least effort is intuitively self-explanatory since the idea behind the principle is simple: all effort should be least. Human beings are governed in their choices and behavior by a universal tendency to reduce effort. Zipf (1949 [1972]) explained the tendency as follows: “each individual will adopt a course of action that will involve the expenditure of *the probably least average of his work* (by definition, *least effort*)” (Zipf 1949 [1972]: 543). The principle of least effort seems to be universal for it has always been intuitively employed in various fields of science. Zipf (1949 [1972]) was the first scholar who acknowledged its existence and operation by giving a proper recognition to least effort. His major contribution consisted in explicit formulation and formalization of the principle. Moreover, he established the proper understanding of least effort, excluding the possible erroneous interpretations. Taken at face value, least effort means the shortest possible way or a shortcut. It must be emphasized that least effort does not mean the simplest. Rather, it is a means to minimize the total effort in the long run. Selecting a longer more complicated and, thus, more effort consuming way may eventually lead to least average of work. This observation constituted one of the most significant achievements of Zipf (1949 [1972]) and ought to be treated on a par with his frequency distribution law.

In order to illustrate the idea of least effort, a metaphor of two towns is discussed. There are two towns of arbitrary names, A and B. It happens that an intervening mountain range separates them. Those people who want to get from A to B can travel through the mountains which would be the shortest way. However, this particular way means that every time the people have to get from A to B, the mountains have to be crossed. Weather conditions may be unfavorable, therefore the shortest way may entail maximum work. Alternatively, the inhabitants can build a tunnel, which involves a huge expense of energy and work. This investment pays off and reduces the workload substantially. The inhabitants make effort in order to build the tunnel but do not need to cross the mountains every time they travel from A to B. Therefore, the

tunnel path ensures least work: “(...) [o]ur selection of path will be determined by the particular dynamic minimum in operation” (Zipf 1949 [1972]: 2).

1.4. Operation of the principle

As far as the stimulus of the principle is concerned, formulation of least effort resulted from numerous empirical observations. Zipf (1949 [1972]) wanted to verify his prediction that our choices and decisions are guided by a common denominator. It turned out that in selecting a course of action, human beings (and phenomena) are motivated by one principle:

We have presented a large number of observations from a truly wide range of living phenomena; this is the empiric aspect of our study, in which we can claim in all modesty to have presented some empiric laws of wide implications. Then each of these different kinds of empiric laws we have attempted to rationalize; this is the analytic aspect of our study. Finally, all these different realizations we have attempted to synthesize in terms of a single unifying principle [...] the Principle of Least Effort (Zipf 1949 [1972]: 543).

The principle of least effort underlies the entirety of human behavior for it concerns individual human beings as well as whole social groups. The principle applies to all aspects of human life and to related domains of human activities. Before Zipf (1949 [1972]), the application of the principle of least effort has been confined to nature or physics, whereas he altered this limiting viewpoint and extended the principle to all manifestations of human behavior. The operation of the principle appears to be relatively uncomplicated and is summarized by the following passage:

A person in solving his immediate problems will view these against the background of his possible future problems, *as estimated by himself*. Moreover, he will strive to solve his problems in such a way to minimize the total work for that he must expend in solving *both* his immediate and probable future problems. That in turn means that the person will strive to minimize the *probable rate of his work-expenditure* (over time). An in doing so he will be minimizing his *effort*, by our definition of effort (Zipf 1949 [1972]: 1).

The principle of least effort is, in turn, governed by the *singleness of the superlative*. This concept, introduced by Edgeworth, is dynamics-related and postulates single use of

any degree-related superlatives.³ In other words, no description can contain two superlatives at the same time if the plural superlatives are in conflict, such as in the case of the superlatives *minimal* and *maximal*. Otherwise, the entity which originally was supposed to be either minimal or maximal becomes minimal and maximal simultaneously. Consequently, the description is vague and devoid of meaning. In order to demonstrate the singleness of the superlative, one can imagine the conditions under which a competition for submarine commander is won: the one who sinks the greatest number of ships will win the prize. The time interval is specified, hence in this case, *the greatest number of ships* is the single superlative. In a slightly altered situation the submarine commander wins the prize for sinking a given number of ships in the shortest possible time, what renders the time limit the single superlative. In both cases the commander knows the conditions for awarding the prize: either the time or the number of ships. If the two separate superlatives are combined into one, that is *to sink the greatest number of ships in the shortest time*, interpretation and comprehension of the conditions becomes severely complicated. The submarine commander now doesn't know the precise terms of the prize since he is at a loss as to whether it is the time or the number of ships that counts. Zipf (1949 [1972]) called this type of situation "completely meaningless and indeterminate" (Zipf 1949 [1972]: 2), pointing to the erroneousness and ludicrousness of statements such as *in a democracy we believe in the greatest number of goods for the greatest number of people*. He argued that one thing cannot be governed by one dynamic minimum and as well as by a totally different one at the same time.

On no account should minimal effort be equated to minimal work due to the fact that the seemingly related concepts of work and effort are not the same. They are used interchangeably just for generalization and convenience sake: "we shall use the term least effort to describe the preceding least average rate of probable work" (Zipf 1949 [1972]: 6. The concept of average rate of work expenditure over time remains a meaningless, vague if not empty notion. Human beings are not capable of predicting future problems at all, let alone predicting what the problems are going to be. Therefore, one can only predict what the problems are likely to be so that one can adjust their behavior accordingly:

³ He is regarded as one of the founders of mathematical statistics and mathematical economics. He also studied applications of mathematical physics, providing insights into, among others, welfare economics, the measure of value, probability and utility.

[B]efore an individual can minimize his of average rate of work expenditure over time, he must first estimate the probable eventualities of his future, and then select a path of least average work through these. Yet in so doing the individual is no longer minimizing an average rate of work, but a probable average rate of work; or he is governed by the principle of the least average rate of probable work (Zipf 1949 [1972]: 6).

Moreover, effort is different from work in that we can estimate effort consciously and manipulate its extent. Using the two towns metaphor, it is possible to calculate the effort necessary to dig a tunnel by assessing costs, materials, labor etc. The effort may be minimized by employing specialist machines or increasing the number of workers. We cannot, unlike in the case of effort, estimate the work to be done without knowing what the work will be in the future. The work of building a tunnel may be complicated by acts of God, labor strikes, oil crisis involving a rise in transport costs, protests of the environment-friendly organizations etc. These complications are impossible to be predicted and taken into account in advance. But when they occur, effort needed to remedy them can be estimated with a relative high probability. As a result of the differences between work and effort, the principle operates along two complementary parameters. The first one is *mentation*, i.e. cognitive comprehension of a problem and probabilities which means that an individual has to calculate and assess the task and effort expenditure. The second parameter is the application of least effort. Having considered the task, an individual employs a way which procures least work. The two parameters are complementary in the way that at first one has to assess the problem, and then one can think how to solve it at least expense of effort: “the work of calculating a path must be included in the total work of taking the path of least effort” (Zipf 1949 [1972]: 11).

The tools-and-jobs analogy reflects the nature of the principle and was drawn in parallel with a means-and-ends, instruments-and-objectives line of illustration. Least effort manifests itself in the two aspects of economy: the economy of tools and the economy of jobs. In terms of relation, tools-seek-jobs and jobs-seek- tools is reciprocal, which can be illustrated with the example of an artisan. He has at his disposal an array of tools and he must perform his job with them so that he minimizes the average probable work rate over time. In the words of Zipf (1949 [1972]), “he must procure an n number of different kinds of matter energy with which he will perform m different kinds of operations upon matter energy so that he can get enough energy to support himself” (Zipf 1949 [1972]: 161). The n variable denotes tools, whereas m stands for

tasks and support energy for jobs. This is a jobs-*seek*-tools case. This situation can be described in terms of a single superlative: minimal as the work to expend is minimal. However, from a different angle this artisan's tools have to perform a job so that the artisan can procure or maintain them, which is the case where tools seek jobs. In conclusion, if an artisan has a job to do, he seeks tools. But if he has tools, he seeks a job. As a matter of fact, the point that end justifies means is crucial for least effort's proper understanding: "[t]he problem, then, is one of altering jobs and tools that he can minimize effort while maintaining production" (Zipf 1949 [1972]: 162). It is not of vital importance whether jobs or tools are used as long as the result, that is least effort, is achieved. An artisan can achieve it by acquiring the target at the minimum expense of energy. In the case of the artisan the principle of least effort will result in the fact that the most frequently used tools will be subject to redesigning with the objective of making it as easy to handle, small and light as possible. In similar vein, the most frequently used tools will lay the closest to an artisan, whereas the greater distance will be characteristics of the least frequently used tools. What is more, an artisan will be prone to working with the nearest tools as reaching the farther ones requires energy expenditure. In the case of speech, the jobs-and-tools analogy becomes slightly altered as an artisan must procure tools, whereas living organism is already equipped with the speech apparatus and has lexicon as the source of tools given a priori, thus, in speech job is crucial.

1.5. Typology of effort: the speaker's and the listener's

The above tools-and-jobs analogy implies that in speech words are the tools whereas the job is successful communication. Zipf (1949 [1972]) viewed speech as a continuum composed of individual words whose objective is to convey meanings. He took into account the two ends of the continuum and, consequently, introduced the typology of effort. It distinguished between the two economies: that of the speaker and that of the listener. Single words, reflecting the speaker's economy, could be referred to as tools. The economy manifests itself in conveying m meanings (jobs) with one word (tools). However, the job of the listener has an opposite or even contradictory purpose since he wants to determine one meaning with n words. The viewpoint of a speaker is called

force of Unification, whereas the viewpoint of an listener bears the name *force of Diversification*: “[f]orces of Unification and Diversification merely describe the two opposite courses of action which from one point of view or the other are alike economical and permissible and which therefore from the combined viewpoint will alike be adopted in compromise” (Zipf 1949 [1972]: 22). Thus, a balance should be struck between the two opposing forces and act as a platform between these two in order to determine both n words and m meanings. Since the goals of the speaker and listener are in conflict, constant trade-offs must take place.

Zipf’s (1949 [1972]) numerous and insightful contributions played a crucial role in advancement and understanding of the principle of least effort. He placed chief burden on formalization of least effort and on the genuine nature of the principle. In this way, he offered an account of least effort which has the potential of providing a framework within which a number of phonological phenomena can be viewed. Furthermore, his works inspired and fuelled an ongoing debate on the weight and role least effort plays in shaping languages. Zipf (1949 [1972]) also brought the principle to general attention and explicitly recognized its existence in all domains of human activity.

1.6. Review of literature

Zipf’s (1949 [1972]) predecessors are listed by Dziubalska-Kolaczyk (2003) in chronological order: Whitney (1878 [1971]), Sweet (1891 [1960]) and Passy (1890), Baudouin de Courtenay (1877 [1972]), Trnka (1936 [1982]), Martinet (1960) and Krug (1998). As far as their theoretical background is concerned, the above mentioned scholars fail to constitute a homogenous group, unlike the following contributors who are affiliated with the Optimality Theory: Silverman (1997), Boersma (1998), Kirchner (1998, 2004), Aylett (2000), Ernestus (2000), Petrova (2001). Nevertheless, the works of all these linguists demonstrate direct or indirect links to the principle in terms of its presence in language change, acquisition and segment inventories.

The principle is so self-explanatory and obvious that it has always been in use in linguistics: “[i]t might not be easy to tell precisely how and by whom the recognition was first made, and by what steps it arrived at distinct formulation” (Whitney 1878

[1971]: 249). Nonetheless, Whitney (1878 [1971]) speculates that in linguistics the principle can be traced back to Bopp (1889 [1974]) who used the principle in language change.⁴ Whitney (1878 [1971]) himself recognized its importance. He viewed language as a natural phenomenon which is subject to least effort and compared the principle to gravity, an unstoppable force that is unidirectional. No lexical item can remain impervious to the influence exerted by the principle: “a relaxation of the tension of effort at any point allows a weakening to slip in (...) the economic tendency threatens everything, and reduces everything whatever is not guarded-or rather, reduces most rapidly what is least guarded” (Whitney 1878 [1971]: 255). The operation of economy is not a “conscious and intended (...) action on the part of the users of a language” (Whitney 1878 [1971]: 254). The principle is opposed by some other principle Whitney (1878 [1971]) never gave name to. However, by analogy, it may be a tendency towards maximization of articulatory effort. This particular issue is open to suggestions as speculations can also lead to a conclusion that the rival principle is constituted by minimization of perceptual effort, for that matter. Furthermore, least effort accounts for the order of language acquisition: “some sounds are harder to catch and reproduce than others” (Whitney 1878 [1971]: 251). Easier sounds (e.g. vowels which involve no air obstruction) are acquired prior to those sounds which require more effort (e.g. affricates, approximants). Least effort also explains language change, in particular, assimilation: “[b]y processes which are completely explainable as results of the tendency of economy, whole classes of sounds are lost from a language or are converted into others” (Whitney 1878 [1971]: 254).

Language is widely believed to have one function: to communicate. In a conversation there is little time to finish an utterance and delivery should be performed as quickly and as effectively as possible. This approach, referred to as functionalist (Sweet 1891 [1960] and Passy 1890), overlaps to a great extent with the principle of least effort. The principle accounts for, among others, loss of unaccented vowels and assimilation processes. These processes lead to language change, ensuring the desired effective and quick realization of particular utterances. The issue of language change was also taken up by Baudouin de Courtenay (1877 [1972]). He viewed the tendency towards convenience as the reason for all linguistic change and speculated that the

⁴ He was a celebrated linguist and a founder of comparative Indo-European linguistics as he demonstrated the relationship between the Indo-European languages.

tendency is one of the general forces shaping the development and structure of language. The force is referred to as a general law “of the development of language” (Baudouin de Courtenay 1877 [1972]: 57) and manifests itself in the process of replacing the difficult sound for easier ones. Least effort also determines the phonetic choices of both the speaker and the listener who wish to minimize their respective workloads in their tasks.

The principle of least effort played a role in phonotactics in the form of general laws of phonetic combinations, proposed by Trnka’s (1936 [1982]). In particular, he addressed the issue whether phoneme combinations within one morpheme are governed by general rules. He formulated the law of the minimal phonological contrast which states that “phonemes differentiated by a mark of correlation never combine in the same morphemes” (Trnka 1936 [1982]: 113). At least two pairs of phonemes which are not connected by another correlative mark constitute the phonological correlation (Trubetzkoy 1931). This law is universal for all languages and allows no exceptions:

Thus we do not find such consonantal combinations as [pf, fp, tθ, θt] and [pb, td, kg], etc. in those languages where the correlations of plosion or voice exist, e.g. in Anglo-Saxon and Modern English. Similarly the combinations [ph - p, p - ph, th - t, t - th], etc. do not occur in Sanskrit and Old Greek, where aspiration is a correlative mark. In German, Czech, Magyar and other languages in which the quantity of vowels is phonological, the combinations of two vowels differentiated only by the correlative mark of quantity (e.g. [ii:], [uu:], etc.) are non – existent (Trnka 1936 [1982]: 114-114).

Another law of the minimal phonological contrast holds that under circumscribed conditions at the morphemes boundary the marked phonemes can be combined with the unmarked ones. In this way minimal phonological contrast is maintained since in the event in which the morpheme boundary ceases to exist, the minimal contrast is reflected as assimilation of the two sounds in question. Thus, phonotactics conspires to produce such clusters which involve minimal effort on the part of the listener in encoding the quality of phonetic combinations.

Linguistic evolution is frequently viewed as the result of constant tension between the conflicting demands of the speaker and the listener. Both the speaker and the listener have certain communicative needs and both are unwilling to allocate more energy than it is absolutely necessary: “[t]he permanent conflict between man’s communicative needs and his tendency to reduce to a minimum his mental and physical activity” (Martinet 1960: 167). For Martinet (1960), the term *least effort* carried the following meaning: “man gives of himself only as much as is necessary to attain the end he has in view” (Martinet 1960: 167). He referred to natural laziness or the tendency to

avoid effort as to *man's inertia*. He also specified the degree to which effort should be avoided. It is the bare minimum of energy that people are willing to expend in performing any activities, speech included. Martinet (1960) advocated the principle of least effort as the key notion to understand the real nature of linguistic behavior. He assumed that the tendency to allocate minimal effort expenditure governs, among others, the preference for selecting shorter and less effort-costly lexical realizations. Accordingly, there are two economies, along with Zipf's (1949 [1972]) proposal: the syntagmatic economy and the paradigmatic one. The former is discussed with the example of replacing proper names with brand ones in which instead of saying *vacuum cleaner* one can say *hoover*. The four syllable name is substituted with a two syllable word, as *hoover* is shorter than *vacuum cleaner*.⁵ The paradigmatic economy means that "we shall avoid a new item in the list of substances which the speaker must memorize and among which he chooses when he speak" (Martinet 1960: 168). Therefore, a choice of a shorter word involves a lesser burden on memory. Unfortunately, energy expenditure in speech is not very well observable because it is balanced by the very urge or wish to utter. The state of physical fatigue allows for the speaker and the listener to comprehend how much energy speaking or listening requires. When we are tired, it is difficult to speak eloquently and find the right wording.

The least effort principle is also present in frequency-related issues. Krug (1998) discussed the findings of an empirical study devoted to the subject of enclitization of function words which provided evidence for the least effort principle: "[t]he present study has provided both support and qualification of this hypothesis that frequency plays a crucial role (...) It suggests that Zipf's proposal has deservedly become tacitly acknowledged received wisdom in linguistics - if indeed it has. If it hasn't, it surely should" (Krug 1998: 311).⁶ Least effort is the underlying principle of his Frequency Factor which corresponds to the speaker's economy. The Frequency Factor is defined as "a force of unification aiming at the minimization of form" (Krug 1998: 300). This factor is counterbalanced by Recoverability Factor which denotes the listener's economy as the listener strives for maximization of informational content. These two

⁵ There exist a few ways to achieve economy, that is substitution (*Kleenex* for *handkerchiefs*) and segment deletion (*Coke* for *Coca Cola*, *Benz* for *Mercedes Benz*).

⁶ He also remarks that Zipf's (1949 [1972]) contribution is underappreciated and quotes enclitization as a vivid manifestation of least effort. Krug (1998) postulates that least effort ought to be ascribed a more decisive role in linguistic behavior.

parameters were modelled on these of Horn (1984). They have antinomic teleologies, but both must be at work in the acts of communication for pragmatic reasons:

If this principle [Frequency Factor] could operate uninhibited, all utterances would result in the sound schwa [...] so that hardly any work would have to be done by the speaker. At the same time, of course, this is the worst-case scenario for the listener, who would have to infer all the meaning from the context. Its counterforce, therefore, [the Recoverability factor], is geared toward the listener's needs and thus aims at the maximization of informational content. It is a force of diversification, and its ultimate goal would be isomorphism (one form corresponds to one meaning). Thus would render pragmatic inferencing, namely, the work on the part of the listener, superfluous (Krug 1998: 300).

However, the use of frequency notion poses a problematic question related to statistics. Statistical information is explained via frequency of use in speech (here enclicitization of function words). Nevertheless, an alternative account can be proposed. Statistics in terms of frequency may be a result rather than the justification of particular use.

The Optimality Theory (henceforth OT) framework has spawned a number of studies in articulatory effort and perception. In particular, the comprehension of least effort in the sense of Zipf (1949 [1972]) appears to be convergent with the search for optimality in language and its choices. In the OT literature and formalism the principle of least effort appears to occupy a prominent position. It is a constraint (MIN EFFORT: Boersma 1998, LAZY: Kirchner 1998), a general principle governing communication or governing specific phonetic realizations (care of articulation: Aylett 2000, reductions: Ernestus 2000). However, the claims that least effort has shaped languages for centuries and via a chain of language changes (obstruents: Petrova 2000, language inventories: Silverman 1997) provide little insight into the true nature of language since such claims already possess the status of basic linguistic facts. It is interesting to note the persistent lack of references to the historical aspect of the principle in terms of past research which gives an erroneous impression that OT scholars have actually discovered the least effort principle.

The most influential approach was that of Boersma (1998) who introduced Functional Phonology within the OT. The work of Passy (1890) provided the stimulus of the functionalist approach. Functionalism in the sense of Boersma (1998) acknowledged the role of minimization of articulatory effort and maximization of acoustic contrast. Specifically, these two principles determine phonological processes and structures which grants them automatically the functional status. The interaction between them gives rise to surface phenomena such as the segment, spreading,

licensing, feature geometry, underspecification and all effects related to the Obligatory Contour Principle (Boersma 1998). Functional Phonology formalized the interaction between articulatory and perceptual drives: “[s]ound systems are not structured directly according to the principles of maximisation of *acoustic contrast* and minimization of *motor effort*, but rather on the minimization of perceptual confusion, which involves acoustic contrast and categorisation, and on the minimization of production effort, which involves motor effort and coordination” (Boersma 1998: 462). Boersma (1998) stressed that categorization and coordination outweigh articulatory effort and acoustic contrast in terms of significance. Thus, he developed MIN EFFORT constraint which penalizes lexical candidates violating an equal number of other constraints of lower rank.

In similar vein, Kirchner (1998) in his study of lenition processes formulated the LAZY constraint on the basis of the principle of least effort: “both the flapping process and the spirantization processes (...) are driven by the same constraint, LAZY (...) and the choice between the spirantized and flapped outputs follow from a single, consistent constraint hierarchy” (Kirchner 1998). In the search of his own lenition approach, he declines the restrictive features inventory claiming that it fails to offer an adequate treatment of phonological contrast. Therefore, he enriched phonological theory with a gradient variation, allowing the abundance of phonetic detail in phonological representation and an adequate characterization of a range of phonological phenomena. The gradient variation of phonological representations was obtained via manipulating the expenditure of articulatory effort as the means of grading phonetic input into representations. Kirchner (1998) proposed an effort based approach to lenition processes: “the class of lenition processes (...) motivates direct reference to a universally non-contrastive phonetic property, namely the articulatory effort (qua biomechanical energy) expended in realizing particular segments” (Kirchner 1998).

In the OT literature, the principle of least effort is responsible for the typology of phonological segments inventories. Silverman (1997) claimed that phonology is a combinatorial system where simultaneous gestures alignment is traded in favor of perceptual recoverability. His thesis was tested in a cross-linguistic environment so that typologically different languages (e.g. Zulu, Comaltapec Chinantec Mpi, and Tamang) are also taken into account. Silverman (1997) found that the phasal relationship between laryngeal and supralaryngeal gestures serves the purpose of maximizing auditory salience. If contrastive values cannot be recovered, gestures are arranged in sequences.

The case of stops aspiration where sequenced laryngeal abduction following the stop closure results in a broadband noise is quoted as an example of gesture sequencing. In the light of the obtained results one can conclude that “there is a typological preference for phasing patterns in which the gestures are optimally recoverable” (Frisch 1998).

The principle was also researched in relation to stochastic suprasegmentals by Aylett (2000). He introduced the notion of care of articulation in order to capture the meaningful differences between a distinct and less distinct speech. He observed that “[i]n general more carefully articulated speech or ‘clear speech’ is longer” (Aylett 2000). Less effort generated a sloppy speech, whereas an increase in articulatory effort resulted in a more distinct speech. He also specified the role of prosody as the platform mediating the demands from both language structure and “the constraint of producing a robust and effective signal” (Aylett 2000). Furthermore, the principle of least effort was identified as the major factor behind reductions in casual Dutch by Ernestus (2000). In particular, she examined obstruent realizations in a corpus-based study of spontaneous speech, observing that obstruents “are realized as voiced when a voiced realization requires less articulatory effort, and as voiceless when it is the voiceless realization that takes the least articulatory effort” (Ernestus 2002). She found that the cases of obstruent final devoicing “are the possible result, in the majority of cases, of the speaker’s natural tendency to reduce articulatory effort, and are partly a matter of lexicon” (Ernestus 2000). Obstruent system was also examined within OT from diachronic perspective. Petrova (2001) investigated language change and concluded that it is the outcome of interplay between motor effort and perceptual salience:

A comprehensive analysis of language change, and, especially, parallel sound shifts, calls for the integration of two complementary approaches within OT: faithfulness (...) and dispersion (...). In the faithfulness framework, language change is viewed as a resolution of the conflict between the tendency to save articulatory effort and the preference for the faithful mapping of input representations to their output correspondents (Petrova 2001).

1.7. Motivation of the principle

The motivation of the principle derives from non-linguistic (physics, psychology and cognitive sciences) and linguistic (low-level phonetic processes and language universals) fields.

Although this notion is rooted in philosophy (Marcus Aurelius 167 AD [1998]), its first application has come from the field of physics: Newton (1687 [1995]) observed that a body, when falling down, performs the movement with the use of minimum energy, just by gravity. However, he concentrated on gravity, neglecting the minimal energy. Next, the physicist Maupertuis (1750) was the first scholar to postulate the principle of least effort in an explicit manner, calling it *least action*. On the basis of his observations, he formulated a law of least action:

[W]hen a mass, M, moves from a given point at a given moment of time to another point at another moment of time, it will proceed along the path in which the sum of all masses when multiplied by their respective distances moved and by their respective velocities will be a minimum (Maupertuis in Zipf 1949[1972]: 13).

Likewise, biology also uses the principle of least effort with reference to animal behavior. This issue was specifically addressed by Gengerelli (1930):

The behavior of an organ elicited by a given stimulating situation which affords relief to an internal need of that organism tends, with repetition, to approach, in terms of time, space and effort involved, the minimal limit compatible with the relief of that need; the nature of the limit being defined by the structure of the organism and of the external situation (Gengerelli in Zipf 1949 [1972]: 14).

Animal behavior was also discussed by Tsai (1932) who observed that it is governed by the principle of least effort: “[a]mong several alternatives of behavior leading to an equivalent satisfaction of some potent organic need, the animal, within the limits of its discriminative ability, tends finally to select that which involves the least expenditure of energy” (Tsai in Zipf 1949 [1972]: 14). Nowadays, the principle continues to apply to animal biology. Palya (1985) conducted a sign-tracking with an interfood clock experiment on a control group of pigeons. This control group was trained to eat when the feeding time was signaled by colourful stimuli. The group learned that the stimuli mean food and responded only when the stimuli period was finished. Least effort was one of the plausible explanations of pigeon’s behavior.

[S]timuli other than the one directly contiguous with food presentation would control chronic sign-tracking. The procedure partitioned a fixed 60-sec interfood interval into 10 6-sec periods, each signaled by a distinctive hue. This "interfood clock" provided a measure of the sign-tracking controlled by each of the 10 stimuli that spanned the interfood interval. It reliably generated and maintained responding to fifth order stimuli. Response rates were successively higher to stimuli that were successively closer to food. The resulting behavior was attributable neither to hue nor to temporal generalization. If responding on all but the final stimulus had eventually ceased, the finding would have

been consistent with traditional notions of least effort, stimulus control, and discrimination (Palya 1985).

Other experimental biologists (Hoyt – Taylor 1981) amassed a large deal of data on energy expenditure used by various species during locomotor tasks. The data demonstrated that animal movements are governed by the criterion of minimum–energy expenditure.

The principle was also taken up in the early twentieth century in psychology. Wheeler (1929) was the first psychologist who introduced the concept into human behavior. Unfortunately, he failed to gather much evidence to support his claim. His line of thinking, however, was continued by others. For instance, Hull (1943) discussed extensively the fact that people avoid effort in general as least effort is a part of their nature: “[i]f two or more behavior sequences, each involving a different amount of work (W), have been equally well reinforced an equal number of times, the organism will gradually learn to choose the less laborious behavior sequence leading to the attainment of the reinforcing state of affairs” (Hull in Zipf 1949 [1972]: 15). Waters (1937) also viewed least effort as grounded in psychology:

Thus Theseus, after slaying the minotaur, found his way out of the labyrinth and to his loved one by following the string which he had carried with him into the labyrinth. Perhaps this was not the most direct route in terms of distance, time, or effort, but it was the only sure way he had of escaping (Waters in Zipf 1949[1972]: 14).

The principle is still in use in modern psychology. For instance, a team of the researchers from Carnegie Mellon’s Center for Cognitive Brain Imaging demonstrated that the human brain employed economy of effort in the process of solving problems (Reichle et al. 2000). The findings suggested that the brain was seeking to reduce the workload substantially, thus selected the strategy which minimizes the workload. There exist two basic strategies which may be applied in dealing with a problem: the visual one which activates the parietal cortex responsible for visual and spatial information processing and the verbal one, which generates brain activation in a network of brain areas, such as Broca’s area, governing speech production ability. The study compared brain activation patterns in thinking of simple tasks. The subjects were instructed on the two strategies and when they were supposed to use each one. In the task they judged a simple sentence *The star is not above the plus (sign)* as true or false in relation to a picture which accompanied the sentence. Comparison of picture content with the

sentence meaning engaged both visual and spatial thinking, whereas comparison of picture meaning with the sentence meaning engaged verbal skills. The results demonstrated that the brains of the individuals who had more skills associated with verbal thinking were less activated in the Broca's area when the verbal strategy was used. Conversely, the brains of the individuals who had less skills associated with verbal thinking were more activated in the Broca's area when the verbal strategy was used. Those individuals whose visual – spatial thinking was developed to a great extent had less activation in the area of parietal cortex when they used the visual strategy. Therefore, activation of either parietal area or the area responsible for speech was lesser or greater relative to which strategy, spatial – visual or verbal was developed in a better way and had more skills associated with (Reichle et al. 2000). The interpretation of the fMRI scans revealed that mental economy underlies the organization of brain areas as the brain evaluates the choice of a more efficient method, here the choice between the visual and the verbal one. The brain performed its thinking service, however it appeared to select a service mode which is less effort costly for an individual. The brain selected this particular service mode in which the individual is more skilled, thus the more efficient mode. The experiment results established the neural basis of strategy and skill in sentence-picture verification and the results indicated that in psychology choice is determined by mental efficiency (Reichle et al. 2000). The results also shed some light on the relation between higher-level thinking and brain-level activity since picture–sentence verification (thinking) seems to be governed by economy of effort in performing the service by brain (activation). Thus, these findings obtained with the aid of brain activation fMRI scans were in accordance with least effort principle.

The principle of least effort is also present in the realm of cognitive sciences. In the hierarchy of economy, principles in cognition and communication govern the constant flow of linguistic information. Fenk and Fenk-Oczlon (1993) empirically verified the Menzerath's Law (Menzerath 1954) which proposes *Sparsamkeitsregel* (the law of economy). It describes the relation between the number of syllables and the number of phonemes in German words. "The interpretations given by Menzerath are aimed at what nowadays is called cognitive economy" (Fenk – Fenk-Oczlon 1993: 11). Fenk and Fenk-Oczlon (1993) tested the Menzerath's Law (1954) in a cross-linguistic study and obtained results in accordance with Menzerath's claim. They concluded that "Menzerath's law serves the 'constant' and 'economic' flow of linguistic information, avoiding an overcharge as well a waste of cognitive resources" (Fenk – Fenk-Oczlon

1993: 23). Their own findings from typologically different languages suggested that there are effective “constraints calling for economy principles in the processing of linguistic information” (Fenk – Fenk-Oczlon 1993: 23).

With respect to linguistics sources of evidence, least effort is present on every single level of the linguistic system: in phonology via the ratio of distinctive features and phonemes, in lexicon via polysemy and the correlations of word length and frequency of token, in syntax via parsimony and simplicity determining the choice of rival structures, in discourse via a balance between costs and effects of any communication act:

[A]ll of these phenomena attests the pervasive influence on language of the Principle of Least Effort. In the final analysis, this principle is grounded in the limits of human life which makes time a scarce resource [...] a language at every point in time affords only the expressive power which is necessary to fulfill the communicative needs of its speech community (Columas 1992: 258-259).

However, the considerations of linguistic sources of evidence in the following sections will be restricted to low-level phonetics and language universals. The former ones are discussed by Ohala (1993):

The study of the phonetic bases of sound change is at the very threshold of being able to make deductive, probabilistic predictions [...] there is a good understanding in phonetics of the mechanisms for turning gestures and postures of the vocal tract into sound [...] there is also some understanding of how physical constraints on this mechanism can give rise to variation (Ohala 1993: 267).

He has been turning to purely phonetic processes in search of explanations. He employed low-level phonetic interpretations in sound change and sound patterns, an approach criticized by Lass (1980) and Dinnsen (1980) for over reliance on pronunciation details. They illustrated the drawbacks of a purely phonetic approach, claiming that it is only capable of accounting for the operational side of a sound change and explicating technical details but fails to account for the reasons why a particular sound change took place. Nevertheless, Ohala (1992) ascribed a probabilistic character to phonetic explication which pertains both to perceptual phonetics and sound change. Ohala’s (1992) approach failed to explain the fact why a particular subject mistook a token of [θ] as [f], however his approach was capable of explaining why 25 per cent of subjects thought that [θ] was [f]. Moreover, he was capable of quantifying points of

similarity. There is a famous claim that it is the listener who is the perpetrator of sound change. In the course of a change the listener misapprehends the function of a phonetic detail, and then the listener may incorporate the erroneous interpretation of phonetic feature into his own phonological system and consistently use it in production as a speaker which may be copied and spread by other speakers. The listener may trigger a change in sound system or not because he can recover the intended form. Otherwise, a sound change results from three different mechanisms: hypocorrection, hypercorrection and correction. These mechanisms entail that the listener fails to recover the intended form (hypocorrection), adds a signal that the sound in question did not contain (hypercorrection) or simply mishears and replaces one phonetic feature with another (correction). Dissimilation which is an example of hypercorrection is defined as “the loss or change of one or more features including whole segment when the whole feature is distinctive at another site within a word” (Ohala 1993: 249). The Latin word *quinque* /kwɪŋkwɛ/ was dissimilated in modern Italian to *cinque* /tʃɪŋkwɛ/, hence according to the dissimilation definition, the feature of lip rounding appears elsewhere than in word initial position in *quinque* and becomes dissimilated. Probably the listener of the Latin word was in doubt about whether the second rounding influenced the first one and under the assumption that labialization was not distinctive, the sound change followed. (Ohala 1993). Ohala (1993) argued that most sounds possess cues which differentiate them from one another. The two syllables /gi/ and /di/ differ in respect of a “sharp peak in burst spectrum around 3 KHz, in other respects the spectra are quite similar “(Ohala 1993: 258). /gi/ can be misheard as /di/ but not otherwise. Likewise, if the listener receives a distorted speech signal and its quality is poor or not clear, a presence or absence of nasal feature serves as the yardstick since vowels are nasalized before nasals (*cut* vs. *cant*):”a knowledge of all the multiple cues used to differentiate words is useful, hints on those multiple cues for a given sound can be obtained by seeing what sort of changes it introduces in neighboring sound” (Ohala 1993: 269).

Any act of communication requires the presence and contribution from two persons, the speaker and the listener (or from one person who is both). They represent a conflict between two dichotomies. On one hand, the speaker is associated with motor effort, thus favors mergers and articulatory ease. The speaker also carries the notions of input and gestures. In production he preserves minimal precision with articulatory implementation at the lowest expenditure of energy. On the other hand, the listener is

associated with acoustic perception. He favors contrast and clarity as well as perceptual specification. The listener also carries the notions of output and feature. In perception he is interested in maximal reproducibility in order to get the signal correctly. Both the speaker and the listener play a role in structuring sound systems. On the other hand, language's function is to convey information as quickly and clearly as possible. These sinequanon demands translate into two criteria which a natural sound system must meet: minimal articulatory effort and maximal distance, which in turn branches into perceptual salience between and within words. For instance, minimum effort and maximum salience combined produce a CV syllable. The presence of a consonant followed by a vowel procures maximum perceptual contrast, as obstruction turns into sonority. Lack of clusters ensures minimal effort, therefore, articulatory effort has an influence on the language segment inventories model and at the same time, auditory salience plays a role in the typology of phonological segment inventories. Other linguistic preference would be to incorporate into phonology the sounds which pose relatively few articulatory difficulties, i.e. plosives because "it is easier to run into a wall than to halt an inch in front of it" (source unknown, in Boersma 1990). If articulation exerted an influence on segment inventories, then the ease of stop production would predict the preference for having stops in all languages of the world, whereas fricatives would occur at a less frequent rate. The data from UPSID on language universals provide evidence for this prediction.

Maddieson (1984) using on the Stanford database, created UCLA Phonological Segment Inventory Database (henceforth UPSID) which initially included 371 languages (Maddieson 1984) and was later extended to 451 languages (Maddieson – Precoda 1989). The updated UPSID includes 451 languages of the world which are compared in equally sized samples. These allow distributional statements of the following type: UPSID recognizes 921 different speech sounds, 625 consonants and 269 vowels and that the maximal number of vowels is 15 (in Norwegian), "i/ u/ a" being the most frequent, whereas among consonants "p/ t /k" are most common.⁷ The data derived from UPSID corroborate the observation that CV structure is preferred: over 70 per cent

⁷ There are also other large-scale segment inventory databases of the UPSID type available, for instance, Stanford Phonology Archiving (SPA) Project (Vihman 1977), initially including 196 languages, is now extended to 209.

of world languages exhibit such preference, or that all world languages have stops” (Maddieson 1984).⁸

The discussion of least effort reflects three major points which should be highlighted. Firstly, the principle of least effort formalized by Zipf (1949 [1972]) has been surfacing in linguistics since its beginnings. Secondly, least effort does not mean the simplest means. A direct, simplistic interpretation of the principle does injustice to its genuine meaning. Thirdly, the quoted instances of research carried out in both linguistic and non-linguistic disciplines furnished evidence for the existence of least effort in various human activities. Moreover, they demonstrated that the operation of least effort is empirically verifiable.

⁸ To be more specific, 70 per cent have no consonant clusters (or less than one) which is very much in accordance with the CV preference.

Chapter Two

Hierarchy and preference in Natural Linguistics: epistemology of the concepts

2.1. The aim of the chapter

The purpose of the chapter is to examine the epistemology of the concepts of *a hierarchy* and *a preference*. First, the chapter presents the theoretical background for the further discussion. Section 2.2. establishes a theoretical framework of the present thesis, i.e. Natural Linguistics (henceforth NL) and outlines the major assumptions of NL. Then, the chapter discusses briefly the concept of a hierarchy. Section 2.3. provides its definitions and reviews its various understandings. Next, the chapter deals with the concept of a preference. Section 2.4. defines the concept in decision theories. Finally, the chapter discusses the concept of a preference in linguistics. Section 2.6. presents preference-based theories within NL.

2.2. Natural Linguistics

Prior to the presentation of the concepts of *a hierarchy* and *a preference*, the theoretical framework of the present thesis must be outlined. In general, one can draw an ideological distinction between the two types of linguistic theories: the formalist theories (e.g. generative theories such as the OT or Government-and-Binding) and the functionalist ones (e.g. NL). These two types differ, among others, with respect to

linguistic complexity.⁹ Simplicity and economy of description and explanation is a priority, however, since language is complex, the complexity of formalisms grows in order to be exhaustive. The adopted formalism (e.g. rules or constraints) is verifiable a priori. If the formalism fails to account for some data, the assumptions are modified. This concerns the strong version of the formalist theories, which in general have a tendency for reductionism and limit the formalism, which otherwise would be capable of explaining everything in principle (Piotr Gąsiorowski, p.c.). The functionalist theories, on the other hand, fully acknowledge the complexity of a language and it seems that in this way they have chosen their fates as they are not able to explain everything in language (Katarzyna Dziubalska-Kołaczyk, p.c.). In search of explanations of language forms, the functionalist theories turn not only to language internal factors (e.g. if the forms result from an interaction between particular components such as phonology, morphology, syntax) but also to the language external ones (e.g. if the speakers of a language alter the forms, if communicative situations select the forms etc.). The ideological distinction between the two types of theories can be illustrated as follows: the vision of OT stipulates that the brain computes candidate forms, whereas the vision of NP is that brain works through processes. The thesis does not provide this comparison in order to evaluate the theories. Both assumptions, as any other assumptions, are equally valid. Summing up the ideological debate, any two theories are comparable only at the point of their basic assumptions, and all debates whose vision is better, make no sense. Neither does comparing explanatory tools. In OT constraints are universal and phonetically motivated. They may be compared to preferences in NP, but OT describes language ad hoc, how the constraints are ranked, whereas NP works out possible scenarios based on predictions, thus the comparison makes no sense. Once the theory selection is made, the theory is discussed internally. One cannot, for example, adopt OT tableaux in NP, because those tableaux derive from the theoretical assumptions (e.g. from constraint ranking).

Therefore, the present thesis adopts Natural Linguistics on the basis of one simple criterion: since the thesis explores a functional principle of least effort, a functionalist theory was selected.

⁹ They also use different sources of evidence (e.g. language internal or external), different approaches to language (e.g. conventional or not), explanatory principles (e.g. processes vs. constraints) etc.

Natural Linguistics is a functionalist theory and continues the line of thinking represented by a variety of researchers of the turn of the 19th and 20th centuries, such as: Sweet (1891 [1960]), Sievers (1893 [1901]), Passy (1890), Jespersen (1921 [1964]), Kruszewski (1883 [1995]), Courtenay (1974), Grammont (1901), Sapir (1921 [1949]) and Jakobson (1941 [1968]). NL as a theoretical framework has been derived directly from Natural Phonology (henceforth NP) which was founded by Stampe and Donegan (1979). Since its inception NP has evolved so much that it is legitimate to call it nowadays *classical NP* as distinct from *Modern Natural Phonology* (henceforth MNP). With the course of time, other linguistic fields have been incorporated into the research. For instance, major contributions in morphology and semiotics came from Dressler (1985), marking the inception of NL. In particular, he elaborated on the concept of naturalness and introduced semiotics as a metatheory to NL. The theory of phonology received formalization and methodology in the form of Beats-and-Binding Phonology (Dziubalska-Kołaczyk 2002a). The works of Dziubalska-Kołaczyk (2001, 2002a, 2002b) represent MNP. For instance, the Beats-and-Binding Phonology (Dziubalska-Kołaczyk 2002a) offered the treatment of numerous phonological phenomena (e.g. processes, representations) without recourse to the syllable, whereas universal phonotactic preferences are derived from sonority distance principle (Dziubalska-Kołaczyk 2001).¹⁰

NL came into existence as a reaction to the structuralist school which highlighted distinctiveness and as a reaction to the generative approaches which stressed the importance of simplicity in terms of rules.¹¹ NL abandoned these approaches and pursued the idea of naturalness. The idea was developed by Stampe (1973 [1979]) and Stampe and Donegan (1979). Under NL, language is viewed as a “natural reflection of the needs, capacities, and world of its users” (Donegan – Stampe 1979: 130). Therefore, humans respond in a natural way to phonetic forces, while the responses are bound by limitations imposed by speech production and perception. The main corollary of the theory is that there is a phonetic motivation behind the structures of the phonological system. The principle on which NL operates is the tension between the two contradictory demands of the speaker (ease of articulation) and the listener (clarity of perception): “(...) natural linguists refer to the tension between contradictory preferences

¹⁰ The syllable in the sense of a basic, non-derived structural unit.

¹¹ Due to the fact that other parts of the thesis deal with NL in a more detailed fashion, the discussion of the theory contained in this section only briefly overviews its most important characteristics.

as the guiding principle according to which linguistic grammars are structured” (Dziubalska-Kořaczyk 2002b: 104). This tension is reflected by the two sets of processes, which serve either the demands of the speaker (lenitions) or of the listener (fortitions, cf. Chapter Five). Due to an emphasis on language users, all statements made by NL have the character of preferences:

(...) generalizing statements formulated in natural linguistics have the status of universal or language-specific *preferences* and not absolute rules or laws. One can gradually move from less to more preferred forms when referring to a preference [...] Natural Linguistics is, thus, explicitly constructed as a *preference theory* rather than a general descriptive theory (Dziubalska-Kořaczyk 2002b:104).

The theory is interdisciplinary in nature and implements the so-called bridge theories with semiotics serving as a potential metatheory: “(...) predictions and explanations are *functionalist* and *semiotic* in nature. One can, to some extent, predict form on the basis of its function; however, a given form may be allowed to serve more than one function, as well as a particular function may be satisfied by multiple forms” (Dziubalska-Kořaczyk 2002b:104). Furthermore, NL makes use of language empiria such as first language acquisition, second language acquisition, aphasia, phonetics, writing systems, psycholinguistics, metaphonology, phonostylistics: “(...) *external* linguistic evidence in Natural Linguistics is regarded as substantive: performance data, such as, e.g., casual speech, speech of young children or speech of second language learners, provides evidence for the structure of the speaker’s competence” (Dziubalska-Kořaczyk 2002b:104). For instance, Natural Phonology in its 1980s shape has already proved effective in researching child speech: “The framework of Natural Phonology is particularly useful in the clinical investigation of children speech disorders as it leads to a developmental assessment of their pronunciation patterns” (Grunwell 1982: 192).

2.3. The concept of a hierarchy

A *hierarchy* is commonly defined as an organization of entities according to the levels they occupy and is associated with a system or an organized structure.¹² Since hierarchy

¹² From Greek: *hieros* meaning *mighty, saint, supernatural* or *hierarchies* meaning *gods-related, of holy origins*. The word *hierarchy* originally served to denote *rule by priests* (Kopalinski 1989).

represents a system at work, the concept is present in a whole range of phenomena and disciplines such as, among others, biology, physics, economy, mathematics, engineering, programming, genealogy, linguistics, psychology, organizations, and the world of business corporations, the church and military institutions. Its prime examples are: Linnean taxonomy, genealogy tree and the Maslov's hierarchy of needs. Alternatively, a hierarchy represents a division of objects into subcategories or subclasses which is "in accordance with an ordering that reflects their complexity" (Audi 1999: 380). Under another definition, a hierarchy "has emerged as part of a movement toward a general science of complexity. Rooted in the work of economist Herbert Simon, chemist Ilya Prigogine, and psychologist Jean Piaget, hierarchy theory focuses upon levels of organization and issues of scale" (Allen – Starr 1982). A hierarchy involves an established order in which the subset entities are centered along one specific criterion. In terms of graphic representation, the concept of hierarchy can be illustrated as a superior and subordinate network of nodes which takes various forms such as a tree, a pyramid, a spectrum or a scale.

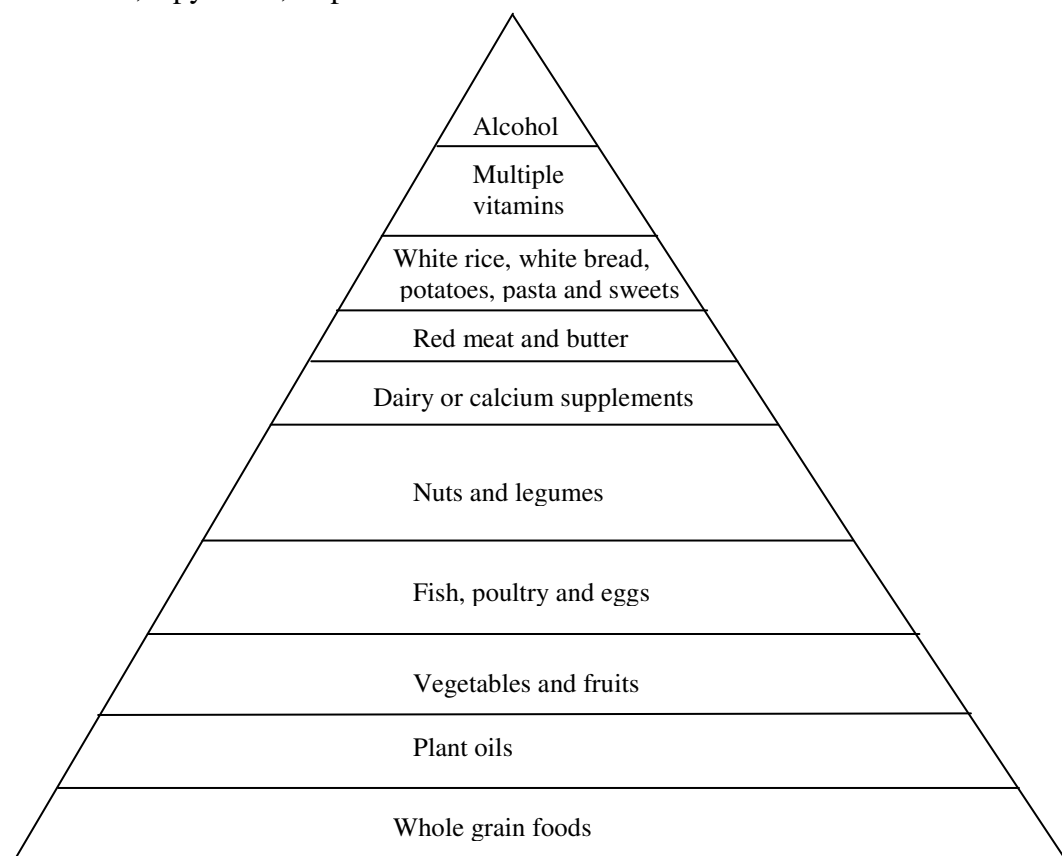


Figure 1. An illustration of a hierarchy - the healthy eating pyramid (after The Harvard School of Public Health 2004).

The operation of the network is relatively uncomplicated due to the fact that hierarchies employ the principle of asymmetrical relationships. In other words, an element of lower rank in a hierarchy can not take over a higher one which results in unidirectionality of the power or importance relationship. This means that in a superior-subordinate situation no role reversal should take place, otherwise the reversal triggers disorders and pathologies (Fritz and Stierlin 1998). As far as the relations between particular entities are concerned, they may be direct, indirect or not occur at all. A number of approaches to the concept of *a hierarchy* will be briefly reviewed. According to Fritz and Stierlin (1998), the concept carries triple meaning. First, a hierarchy describes the structure and function of an authority which is exemplified by a generation gap or parent-child relations. Second, a hierarchy refers to a logical organization in which the order of lower elements constitutes a higher order. The third meaning comprises the order of levels within the system, from the lowest to the highest ones. The latter meaning is crucial for the present thesis. A hierarchy of preferences will be examined in relation to the third meaning. Accordingly, the system levels will be ordered on a scale.

A hierarchy is subject to typology. There are a few types of a hierarchy, the most frequently employed types are nested and non-nested hierarchies. A group of soldiers in an army illustrates the type of a nested hierarchy. The group is a part of a larger system on which it depends, whereas the food chain represents a non-nested hierarchy as a system on its own, independent of a larger structure: “[n]ested hierarchies involve levels which consist of, and contain, lower levels. Non-nested hierarchies are more general in that the requirement of containment of lower levels is relaxed” (Allen, nd). Less frequently used for any system representation is *tangled hierarchy* whose operation appears to resemble a loop. A tangled hierarchy can be illustrated as follows: if one moves oneself through the subsequent levels of a system, one finds oneself in the primary position and comes through the system to the original starting point (Hofstadter 1979).

As far as linguistics is concerned, the concept of a hierarchy is applicable to language due to the features of language. Language discreteness, on the one hand, means that an utterance is divisible into units such as texts, utterances, sentences, phrases, words, moras, syllables, segments (consonants, vowels), and distinctive features. On the other hand, language is hierarchical or rule-governed, i.e. one unit is made of several lower ones. Therefore, a linguistic hierarchy is “a classification of linguistic units into a series of successively subordinate levels, especially an analysis of

sentences into clauses, phrases and words, and morphemes” (Crystal 1995: 453). Trask (1996) explained a linguistic hierarchy as “any of various linear scales among which certain phonetic or phonological elements are ranked with respect to certain properties” (Trask 1996: 170-171). He quoted the examples of the Sonority Hierarchy and the Animacy Hierarchy.¹³ The following definition and illustration of the phonological hierarchy can be found on the website of the Summer Institute of Linguistics: “[t]he phonological hierarchy is a ranking that organizes a stream of speech into levels of ascending size and complexity (...) Units at one level of the hierarchy cluster together to form units of the next higher level. At each level, speech can be segmented into units that have similar kinds of phonetic features”.

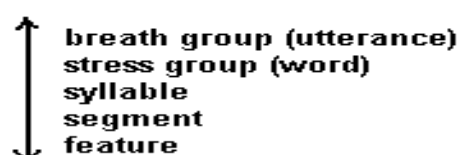


Figure 2. An illustration of the phonological hierarchy (after SIL International 1999).

A containment hierarchy of classes of formal grammars has been proposed by Chomsky (1956). These classes, in turn, generate formal languages. He postulates that a formal grammar consists of a finite set that contains terminal and non-terminal symbols, arranged according to a finite set of production rules.

2.4. The concept of a preference

A preference (from French *préférence* meaning *transfer, conveyance*) is commonly referred to a choice, real or hypothetical, between alternatives. These alternatives can be rank ordered. The concept of a preference is used in social sciences, particularly in economics. The simplest understanding of a preference is as follows: a simple majority X defeats Y when the number of voters who prefer X to Y is greater than the number of voters who preferred Y to X (Sen 1964).¹⁴ He argued that not only the number of voters

¹³ Animacy Hierarchy determines the order of NPs. It stipulates that first and second person outrank the third one, pronouns outrank common nouns, human nouns outrank nonhuman animate nouns, nonhuman animate nouns outrank inanimate nouns.

¹⁴ He is the 1998 Nobel Prize Laureate in Economics.

can determine a preference, but also how much each group prefers one alternative to the other. This statement introduced degrees into the concept of preferences. If 48 per cent of subjects prefer apples and 52 per cent prefer oranges, preferences can be barely inferred, or with a low statistical significance. If the ratio is 20 per cent (apples) to 80 per cent (oranges), one can declare that the preference for oranges to be four times stronger. The above generalization can be specified with respect to the human factor in preferences: “opinion isn’t to be quantified, but rather weighed” (Morales 1797). A preference implies the existence of two opposites: preferred and dispreferred. There can be a wide spectrum of less or more preferred between them which entails that preferences are invariably organized along a hierarchy. If A is preferred to B, there must be both A and B and the relation between them must be also determined. One of them is higher or lower on the scale. The concept of *preferred* evokes the notions of *better* and *more frequent* (Hansson 1984) and, consequently, a better choice is always preferred, and, in turn, more popular in terms of distribution: “[w]orse is the converse of better, and any verbal idiosyncrasies must be disregarded” (Brogan 1919: 97). A preference overlaps with markedness: “[t]he unmarked preferred option (...) has a less complex structure than the unmarked, non-preferred option” (Bussmann 1996: 377). Relative/scaled markedness in preferences means that X is preferred to Y because X is simpler and Y is more complex.

The notion of preference is employed in microeconomics in the form of decision theories. In these theories, preference relations model entities, consumer goods or consumer preferences. In formal language, if a consumer chooses goods A over the goods B, then a relation holds between the goods. This relation is called a *preference*. It can be represented with mathematical notation: $A > B$ which means that A is better than B. In the formal language of preference logics, the comparative notion $>$ describes a *strong* or *strict preference*. If a consumer equally values goods A and B, this preference can be represented as $A = B$. The notion \geq describes a *weak preference*. If a consumer considers goods A to be at least good as goods B, this preference can be represented as $A \geq B$. The notion \geq describes *indifference*. There are two rules governing the three notions (Hansson 1994):

- (1) $A > B$ if and only if $A \geq B$ and not $B \geq A$ (A is better than B if and only if A is at least as good as B but B is not at least as good as A).
- (2) $A = B$ if and only if $A \geq B$ and $B \geq A$ (A is equally good as B if and only if A is at least as good as B and also B at least as good as A).

The notions which describe particular preference relations are comparative and binary as they relate two entities or arguments with each other. These notions possess a number of formal properties, of which *completeness/connectedness* and *transitivity* are used in the decision theories.¹⁵ First, *completeness* is discussed. Any preference relation applies to a finite set of entities for which the relation holds. In the example discussed above, consumer goods A and B constitute the set. If fruit are the set, it is illogical to compare bananas to books which belong to a complete different set. A preference relation has fruit as its *domain* (not books). The completeness property is defined for a preference relation and its domain: The relation \geq is complete if and only if for any elements A and B of its domain, either $A \geq B$ or $B \geq A$. *Transitivity* guides a decision and a consumer can transfer his preferences accordingly. In a transitive preference, a consumer has apples but prefers bananas to apples and oranges to bananas, he is ready to pay one dollar to trade apples for bananas, then one more dollar to trade bananas for oranges. The preference becomes intransitive if the consumer pays one dollar to trade oranges for apples. A strong/strict preference relation $>$ is *transitive* if and only if it holds for all elements A, B, and C of its domain that if $A > B$ and $B > C$, then $A > C$. A weak preference relation \geq is *transitive* if and only if it holds for all elements A, B, and C of its domain that if $A \geq B$ and $B \geq C$, then $A \geq C$. An indifference relation $=$ is *transitive* if and only if it holds for all elements A, B, and C of its domain that if $A = B$ and $B = C$, then $A = C$. If a preference is both complete and transitive, it is a rational preference relation.

In group decision making, for a given class of people and a given class of objects there is a given preference profile. The task for a researcher is to ascribe a common hierarchy of preferences to a given profile (Szaniawski 1994). Game theories use the parameter of preference intensiveness which allows taking into account an interval scale which specifies the relation between degree differences in particular preferences. The preference intensiveness is an overarching parameter. The order of intensiveness of preferred objects on the interval scale implies their preference order, but never vice versa. It is impossible to infer from the less or more preferred order what the details are and by how much things are preferred or dispreferred from others. It is possible, however, to conclude the difference in preferences by means of judging the

¹⁵ Behavioral economics has been developed in order to verify the issue whether the behavior of consumers is consistent or not with these two properties.

interval scale that makes possible to perceive minute differences in differences that Szaniawski (1994) refers to as “stosunku różnic między stopniami ocen” [the ratio of differences the between grades degrees, translation mine, MK] (Szaniawski 1994: 464).

2.5. Preference-based theories in linguistics

This section presents the preference-based theories developed by Stampe (1973), Dressler (1997), Vennemann (1983, 1988) and Dziubalska-Kołaczyk (2001). The aim of the presentation is to examine the use of the concept of *a preference* within NL theories.

2.5.1. Stampe (1973)

All naturalist theories stipulate that preferences are based upon language acquisition and language use. In NP, actually, it was Stampe (1973) who introduced preferences in from of consequences for a language before Dressler (cf. Figure 3). The functionalist principle of naturalness predicts that due to the contradictory interests of language users, speaker and listener, preferences fall accordingly into the articulatory and perceptual sets (Stampe 1973). The articulatory preferences are governed by the need to achieve simplicity, realizing this goal by means of assimilations and reductions (speaker-friendly processes). Therefore, a cluster of segments tends to be simplified in order to facilitate articulation. These processes are particularly conspicuous in languages which have clusters consisting of many consonants. For instance, a Polish word *wszystko* ‘everything’ is reduced in casual speech to /ʃɪskθ/ (the two clusters are simplified). In English, the articulatory preference for similarity manifests itself in all kinds of assimilation processes (e.g. *bad girl* /bæd gɜːrl/ becomes /bæg gɜːrl/) as the speaker prefers to have similar sounds (two velars in the immediate vicinity) in a sequence. Thus, speaker prefers the processes which result in cluster simplicity and similarity of place/manner of articulation, whereas listener prefers processes maximizing perception (Stampe 1973).

2.5.2. Dressler (1997)

The semiotic principle of naturalness implicitly uses the concept of a *preference*, e.g. there exists a well observable preference for iconicity (Dressler 1997). Prior to the discussion of preferences in terms of semiotic parameters, the proper understanding of the concept in linguistics must be established. Preferences cannot be identified with inductive statements which are generated by statistics. Statistically, there are more right-handed people than the left-handed ones. Thus, the conclusion that right-handedness is preferred is circular since there are more right-handed people, there is a preference for right-handedness (Dressler 1997). Rather, preferences, governed by higher principles, have consequences in linguistics, which manifest themselves in a way specific for a language. For instance, the least effort principle (a higher, universal principle) predicts that devoicing of final obstruents is preferred than non-devoicing (linguistic preference). The German language users do devoice obstruents in word final position, whereas the users of English do not (specific, concrete consequence of preferences).

higher principles (e.g., the principle of the least effort, of cognitive economy)	<i>non-linguistic</i> (<i>cognitive, phonetic, psychological, sociological etc</i>)
preferences (e.g., a preference for simple phonotactics, for a CV structure)	<i>linguistic</i>
preference parameters (pronounceability, perceptibility)	<i>functional and semiotic</i>
consequences of preferences (absence of clusters in a language)	<i>linguistic</i>

Figure 3. The explanatory system of Natural Linguistics (after Dziubalska-Kołaczyk 2002b:104).

Thus, preferences govern choices, whereas choices never govern preferences. This particular line of thinking about preferences derives from the Aristotle's notion of *preferability* (haire 'to teron – 'the worthier of choice') which has been, apart from

linguistics, employed in economics, psychology and philosophy of values, psychological personality research, psychological gestalt theory relating to perception and recognition as well as game theory and decision theory (both economical and psychological). These non-linguistic, preference-based theories face an important task: since a preference is structured like a simple majority and A is preferred to B, certain criterion for this preference must be selected and adopted. Moreover, there can be more than one criterion, which means that the criteria need to be arranged on a scale or a hierarchy according to some order. They must not enjoy an equal status, otherwise a preference relation does not hold. Furthermore, preferences in these non-linguistic theories are differentiated into intrinsic preferences (in which betterness is intrinsic and external circumstances are not taken into consideration), normal preferences (established by default), probabilistic preferences and relative preferences. Preferences are also subject to parameters such as transitivity (if A is preferred to B and B to C, then A is preferred to C) and asymmetry/irreflexivity (if A is preferred to B, then B is not preferred to A). If a conflict between preferences arises, it should be resolved to the benefit of these preferences “strive towards maximal benefits or expected utility” (Dressler 1997: 392). Unfortunately, unlike non-linguistic theories, linguistics does not extensively employ the concept of a *preference*, probably due to the fact that it is grossly underappreciated in general. It implicitly exists in markedness theories, invisible-hand theories of language change and optimality theories, whereas explicitly in Vennemann’s (1998) treatment of sound change and in naturalness theories.

The latter theories, however, incorporate preferences in the form of a subtheory of universal markedness which is defined by an intrinsic, universal preference for iconicity parameter: “[w]ithin this subtheory, universal preferences are established and explained by their non-linguistic foundations, e.g. the preference for iconicity (...) from the contributions of iconicity to processing ease in production, monitoring and perception” (Dressler 1997: 395). The iconicity parameter derives from semiotics and, as any other preference, must meet the following criteria: asymmetry, transitivity and connexity (Peirce 1991). These subparameters are illustrated with an example from morphology. English-speaking children in the course of language acquisition are reported to derive frequently the plural form *foots* from *foot*. This form is extremely iconic, thus natural, unlike the adult, correct plural form *feet*. According to the asymmetry subparameter, if A is preferred to B, then B is not preferred to A. Consequently, if the form *foots* is more iconic than *feet*, then *feet* are less iconic. The

transitivity subparameter allows for the following scale: *foots* is on its top in terms of iconicity, then *feet*, next *fish* (non-iconic) and on the other end of the spectrum (anti-iconic) there is *hon* (the plural of Franconian *hond*). The third subparameter, connexity, means that there is the continuous reduction of iconicity: diagram-metaphor-non-iconicity-anti-iconicity (Dressler 1997). The application of these three subparameters of iconicity demonstrates that the plural form *feets* which exists in child language is more iconic, and thus, preferred. The order of language acquisition corroborates this preference as children spontaneously add the plural suffix –s to all nouns as an overregularization and only at later stages of development, by imitation, they acquire the correct form (here *feet*). The same observation in terms of a preference for iconicity holds true for irregular verbs as children are reported to prefer the form *goed*. This particular preference for the iconicity parameter is applied under certain conditions since it results from an interaction between phonology and morphology. Preference for iconicity may also result from local pragmatic conditioning since local conditions “may affect a decrease of markedness (unnaturalness)” (Dressler 1997: 399). The preference applies to conditions in which the parameters interact, as in the case of the *veni vidi vici* dictum. This dictum has the following iconic relations: diagrammatic *ordo naturalis* (the order of events described by the dictum is precisely reflected), indexical iconicity (in repetitions as far as syntax, morphology, alliteration and prosody are concerned) and diagrammaticity (the relationship between Caesar’s rapid victory and the length of the compact phrase as one verb corresponds to one sentence).

Therefore, iconicity interacts in this dictum with the following parameters: first, indexicality in terms of unmarked, anaphoric indexicality on the text level since iconic parallelism compensates for the lack of anaphors. Second, the number of members of a sign relation which means that binary relations are preferred over the unitary, ternary, quaternary ones and so forth. In *veni vidi vici* all the words are bisyllabic. Third, the parameter of figure and ground where the emphasis is on the figure and not on the ground. In the dictum, only the major points, or rather the results of actions of the story are highlighted, without presenting the details: “in narratives, the main event line and the main character receive more attention and coding than marginal lines and secondary texts” (Dressler 1997: 404). Fourth, the parameter of transparency which is preferred over opacity. In the discussed example the text directly reflects the cognitive, semantic or pragmatic meanings. These preferences may interact and combine their effects. Nonetheless, they may come into conflict which must be resolved by setting up higher-

order preferences, such as the conflict between iconicity and binarity. In French, the sound made by a chick is *bi bi bi*. On the basis of the sound, iconicity should conspire to produce a noun *bi bi bi* but in fact, the noun has the form *le bi bi*. The case of the noun *le bi bi* demonstrates that locally binarity is preferred over iconicity.

Dressler (1997) reports on a subtheory of type adequacy which addresses the issue of non-universal cross-linguistic preferences for universal preferences. The subtheory predicts that the agglutinating language type shows a higher preference for iconicity of construction and also for indexicality and transparency than the inflecting-fusional type. In inflectional morphology, language-specific system adequacy refers to "what is normal in a given language (...) what fits normalcy best. Thus system adequacy implies conventional (i.e. symbolic) elaboration on, and specification of, universal and typological preferences, and, even more than type adequacy, it may reverse universal preferences and thus lead to markedness reversal" (Dressler 1997: 406). However, universal preferences "should be more effective in the absence or near absence of system adequacy" (Dressler 1997: 406). This is the case of reduplications, blends and back-formations which belong to the realm of extragrammatical phenomena in morphology. Also, in text linguistics *ordo naturalis* maps linguistic expression iconically and is present in all languages (for instance, in recipes).

2.5.3. Vennemann (1983, 1988)

Another contribution to the discussion of the *preference* concept comes from Vennemann (1983, 1988). He has developed a theory of preferences which accounts for syllable structure and its change. One of the preferences is that open syllables are better than the closed ones. This, in turn, implies that having only open syllables is preferred to not having open syllables. On the basis of the Universal Consonantal Strength, Vennemann (1988) formulated the following laws for individual syllables: the Head Law, The Coda Law, The Nucleus Law, and for sequences of syllables The weight Law, The Law of Initials, The Law of Finals, The Strength Assimilation Law and The Contact Law. He specified that this list is not at all exhaustive and implies the existence of Shell, Body and Rhyme Laws. In formulating his laws, Vennemann (1993, 1998) departed from the standard linguistic classification, in which, typically, processes and phenomena observable in linguistics fall into one of the two categories:

good/natural/unmarked or bad/unnatural/marked which is a descriptive way of dealing with language data and, consequently, invited the use of external evidence: “[m]y conception of preference laws differs from the most approaches to linguistic naturalness by characterizing linguistic structure (...) as better or worse” (Vennemann 1988: 1).

On the basis of language external evidence, he formulated preference laws and introduced graded preferences: “a gradual concept of linguistic quality relative to a given parameter” (Vennemann 1988: 1), which entails that a change of a certain parameter involves the change of the situation as a whole. Under some circumstances parameter A may be better, i.e. more preferred, but under different circumstances the same parameter A can be worse (less preferred). Therefore, the relativity pertains to the parameter’s circumstances: “nothing in this world is good or bad *an sich*” (Vennemann 1988: 1). The concept of graded preferences gives rise to a hypothesis that the less preferred structures have a tendency to change. The less preferred a structure is, the more it is prone to modification and improvement. The notion *change* means for Vennemann (1993, 1998) *a change for better*. Therefore, bad structures change in order to improve and to be more preferred. The hypothesis is verified against the assumption of locality in language change: “every change in a language system is a local improvement” (Vennemann 1988: 1). Locality triggers a chain-like reaction: if one parameter is modified and this is a change for the better, then some other parameter can suffer a different fate and get worse. Locality means also that a linguistic change may improve and deteriorate a structure at the same time. Improvement derives from application of preference laws. Deterioration, however, can result directly from improvement of one of the parameters which leads to the deterioration of another. The processes of diphthongization and copations, i.e. syncope and apocope serve as an illustration of such deterioration: “both (...) by necessity change the syllable structure of the affected items or even the language system, and indeed, always worsen it” (Vennemann 1988: 2). Diphthongs turn the optimal, universally favored CV structure into a less preferred one, VV. Apocope and syncope change the optimal pattern CV.CV.CV into CV.CVC or CV.CCV.

On the basis of the hypothesis that the less preferred structures have a tendency to change, the Diachronic and Synchronic Maxims are formulated. The former one stipulates that “[l]inguistic change on a given parameter does not affect a language structure as long as there exist structures in the language that are less preferred in terms of the relevant preference law” (Vennemann 1988: 2). This maxim governs any

language change and means that the worst structures are the first targets of change on a given parameter. The former one postulates that “a language system will in general not contain a structure on a given parameter without containing those structures constructible with the means of the system that are more preferred in terms of the relevant preference law” (Vennemann 1988: 3). In a normal situation, the range of structures is subject to change on a given parameter. The result of changing from less into more preferred structures is a greater number of more preferred structures. They will gradually eliminate the worse ones and the range of total structures will shrink, thus, the structures range will be smaller in size due to the fact that the less preferred structures are absent from the system. However, changes do not invariably proceed in a smooth fashion. Parallel parameter changes, borrowings, natural gaps due to combinatory constraints may create discontinuous ranges of structures. Still, these would be extraordinary cases.

Preference laws are universal and have phonetic motivation in terms of production and perception: “[t]hey are theorems of a general theory of human communicative capacity” (Vennemann 1988: 4). However, linguistic preference theories cannot explain universals of language because they are language specific (Vennemann 1983). Certain language-specific preferences may clash with the universal preference laws. This is so because languages are not natural, but rather “products of human history” (Vennemann 1988: 4). Although Vennemann (1988) claimed that structures are derived from language historical development, humans approach them with a natural endowment. This observation is corroborated by the existence of language universals and sharing common structures among various language systems. In applying preference laws, one must be aware of the fact that “there can exist no optimal language system as such, but only systems that are optimized on some parameters” (Vennemann 1988: 65). The optimal syllable structure, CV, is ideal only when considered in isolation. The strength of its onset, however, makes the CV structure a bad one when it attaches to other syllables. In Italian, the word *aera* ‘air’ changed to *ai.re* and then to *a.ria*. In *a.ria*, the consonant /r/ is stronger than the vowel /i/. Thus, the metatethetic change failed to improve the syllabic structure. The Contact Law stipulates that sonority sequence is crucial - in onset it must rise, whereas in coda it must fall. Accordingly, the syllable structure is improved when it meets the order of the sequence. This is the case in Spanish, where the bad structure *bas.io* ‘a kiss’ evolved into a better one, *bai.so* and finally, into *be.so*. Vennemann (1988) determined the

explanatory power of his preference laws. If we know which syllable structure is preferred, we can gain a better understanding of historical changes and account for the fact that a certain structure turned into a more preferred one. Preference laws guide us through the history of language as well as synchronic universals: “knowing them we do not stand in the dark vis a vis individual syllable structure restrictions, but understand them - we recognize their motivation. The preferences are rooted in human phonetic endowment” (Vennemann 1988: 67).

2.5.4. Dziubalska-Kołaczyk (2001)

Dziubalska-Kołaczyk (2001) contributed another approach to preference-based theories. For her, a preference implies a human agent and reflects the speaker’s strategies. It roughly equals to markedness: the more preferred means the less marked. The preferences in use and acquisition of language become frozen in that language’s structure (Dressler 1997). She has developed a model for phonotactics in which the phonotactic constraints on consonant clusters are a case of universal preferences. It is estimated that 70 per cent of the world’s languages have no consonant clusters (Maddieson 1984). Those languages which have them are organized in phonotactic terms. The organization follows phonotactic preferences. In a cluster space, they specify a hierarchy of clusters: the preferred clusters which are possible, dispreferred clusters which are possible but dysfunctional in a position and impossible clusters. Phonotactic preferences serve two functions: to counteract the CV preference and to counteract the dysfunctional clusters. The phonotactic preferences are determined on the basis of the Optimal Sonority Distance Principle (henceforth OSDP). OSDP predicts that the most preferred clusters are those which show the greatest difference between values of *sondis* (sonority distance) and allows a conclusion that the clusters preferred in a given position in a word are not preferred in other positions. Dziubalska-Kołaczyk (2001) develops the universal cluster space which is a matrix against which we can compare language-specific phonotactics. Universal phonotactic preferences (perceptually based) become overridden by some other preferences such as articulatory easy phonological sequences and morphological preferences.

The discussion of the concept of *a hierarchy* and *a preference* explains the use of the two concepts in the present thesis. A hierarchy will be treated as the order of levels within the system, from the lowest to the highest ones, whereas preferences will

be specific, linguistic consequences of higher level principles. It also demonstrates that preferences are employed extensively within NL theories and their parameters have been discussed at length. Finally, the discussion clearly determines that NL is the theoretical framework for the thesis due to, among others, its functionalism, the use of external evidence and the fact that the principle of least effort is recognized explicitly, enjoying the status of a higher-level principle.

Chapter Three

Articulatory effort

3.1. The aim of the chapter

The aim of the chapter is to consider the articulatory aspect of effort. First, the chapter reviews the notion of easy and difficult sounds. Section 3.2. provides selected views on the notion. Then, the chapter discusses the concept of *articulatory effort*. Section 3.3. overviews various understandings of the concept. Finally, the chapter presents the concept of *articulatory gestures*. Section 3.4. places the concept within the framework of Articulatory Phonology.

3.2. The notions of easy and difficult sounds

Since the concept invariably evokes the notions of *easy* and *difficult* sounds, prior to presentation of the concept, the notions of *easy* and *difficult* sounds should be discussed. It must be stressed, however, that the present chapter discusses sounds difficulty exclusively, whereas the related notion of ease of articulation is discussed in Chapter Six (cf. 6.5.).¹⁶

The concept of articulatory effort itself is simple to understand, but difficult to apply. In fact, the understanding of articulatory effort is tacit and has never been

¹⁶ The notions of easy or difficult sounds seem to be used paradigmatically, with reference to the sounds of a system whereas the notion ease of articulation is evoked in phonological processes which take context into account, i.e. are used syntagmatically (cf. 6.6).

determined in a straightforward way. Partly due to the fact that it resists to be defined, and partly because it is extremely intuitive and self-explanatory. By the same token, certain sounds or even sound classes such as *obstruent* and *sonorant* differ in effort expenditure, being termed as less or more difficult in production. Therefore, the concept is inseparable from the notions *easy* and *difficult*. Remarkably, the existence of more and less difficult sounds to production had been noticed long time ago: Plato (360 BC [1991]) made an observation that certain sounds require greater effort expenditure than others, claiming that in articulation of /r/ the tongue assumes a strange position, as if it was *agitated*, therefore, its production is energy-consuming (Plato 360 BC [1991]).¹⁷ His observation reflected naive feelings about language, expressed by its users.

Thus, the concept of articulatory effort implies the existence of easy or difficult sound in articulation. Articulation is governed by three parameters: supralaryngeal gesture which accounts as a major factor, speech rate and muscular tension (Busmann 1996). More specifically, if one adopts the understanding of articulation as a sequence of muscle movements, it becomes evident that certain sounds involve more muscle movements or gestures than others, that is to say, aspirated stops are more costly to produce because they involve one articulatory gesture more than the unaspirated stops. In the case of vowels, schwa is the easiest to produce as this vowel entails maximal inertia. Accordingly, a hierarchy of sounds difficulty could be established. However, this line of argument is not held generally since there are two perspectives which view sounds difficulty in different ways: the cross linguistic and the intralinguistic one.

First, a cross-linguistic perspective values sounds difficulty in relation to various language system. A native speaker in the course of his or her first language acquisition accommodates articulators via training to the mother tongue's phoneme inventory. Thus, articulation of, say, velar nasal poses no difficulties for an English speaker. However, it will in the case of a person whose native language lacks this phoneme. For the latter speaker, velar nasal is labeled as *difficult*, whereas for the former one the /ŋ/ sound fall into the *easy* category. Thus, difficulty in the cross-linguistic perspective is a matter of training. Martinet (1960) claims that it is not possible to measure individual sounds without any comparison or other sounds to be valued against:

¹⁷ It is interesting to note that /r/ is reported to cause the greatest difficulties for children and is acquired as one of the last sounds (Slobin 1985a, 1985b).

As regards phonemes, which interest us directly, how could we measure and compare the average energy required for the pronunciation of [a] and [f]? We could at best suppose that the pronunciation of /main/ requires on an average more energy than that of /mai/, that is, other things being equal, an extra phoneme requires an extra amount of energy (Martinet 1960: 172).¹⁸

In similar vein, Ladefoged (1990) made the observation that no sounds are easy or difficult *per se*. He was of the opinion that statements of difficulty made regardless of the observer are not valid. He claimed that for a Navaho ejectives come easier than dental fricatives. This observation takes into account exclusively the cross-linguistic perspective as for the same Navajo, stops come easier than ejectives. Pouplier (2003) addressed the issue of articulatory effort and presumed that any act of articulation involves the expenditure of effort as effort is inherent in any act of communication *a priori*. This particular argument fails to allow the existence of gradability which is capable of establishing the differences between, for instance, aspirated and unaspirated stops. Moreover, Pouplier (2003) continued the line of thinking represented by Martinet and Ladefoged: "There is no context-free statement of effort. Neither a particular speed, nor a particular displacement, nor a particular gesture is *per se* effortful" (Pouplier 2003). Thus, in relation to *easy* and *difficult* sounds, numerous linguists complement the notion *difficult* with the question *for whom* unless one considers the intralinguistic perspective.

The second perspective stipulates that the value of individual sounds can be established against values of sounds representing different classes within one language. Maddieson (p.c.) argues that segment difficulty is a legitimate claim, which should be considered on an intuitive level (e.g. precision) and an articulatory level. It also must be assessed within one language. The existence of an intrinsic scale of difficulty in the sense of sounds of the same language can be supported on two accounts, i.e. linguistic universals and child language acquisition. In terms of language universals, there exists a correlation between phonemes difficulty and token: the easiest sounds are the more frequent. According to Maddieson (1984), all world languages have stops from which one can infer that they are the easiest sounds to articulate. Thus, every single phoneme inventory includes stops, whereas only few languages have nasals. There is, however, a trap in suggesting that frequency explains difficulty. Frequency-related observations make use of statistical data which has always been triggering the problem of circularity.

¹⁸ The stipulation "other things being equal" makes the difference here. In emphatic speech as compared with the casual one things are not equal anymore.

The question remains whether sound difficulty ranking on the basis of frequency is circular or not. There are two angles from which difficulty and frequency are viewed: either sounds are difficult because they are rare or they are rare because they are difficult. For instance, rare velar nasal may be in fact easier to produce than a stop since velum drops by gravity. In terms of child language acquisition, one could possibly establish a scale of difficulty by arranging the sounds according to the order of acquisition. This procedure stipulates the between three dimensions of difficulty: 1) physical production (articulatory distance, coordination and precision) 2) perception: relevant in language acquisition (although children are visually aid and they can map labial gestures or tongue protrusion, velars and uvulars disallow visual mapping) 3) system difficulty (segment difficulty in comparison to that of other sounds of that system) (Ian Maddieson p.c.).

For Maddieson (2005), effort is a correlate of exoticness which he defines as rarity. Clicks are the most exotic segments, they occur only in languages spoken in certain parts of Africa. However, speakers of many languages are familiar with clicks since they employ them to signal emotions (as a sign or disapproval). The difference in perception lies in the fact that in the African languages click belong to phoneme inventory, whereas in English they have paralinguistic function. In phonetics, segments themselves are not exotic but segment sequences or contrast number might be low in occurrence. Rarity is also linked to difficulties in production and perception and explained via frequency: if X is rare, it must be difficult. Maddieson (2005) treats rarity as different from complexity. Being rare involves lack of token, whereas complexity fails to determine the distribution (Lindblom – Maddieson. 1988). Next, the facts of language acquisition give evidence of regularities underlying the acquisition order. For instance, acquisition stages are observed on the basis of sound inventory development. If the intrinsic ease of sounds were equal, then the child would acquire all sounds at once because they would be treated equally by his or her phonology. Language acquisition order appears to corroborate the difficulty ranking: liquids are acquired late by children (Slobin 1985a, 1985b), they are also not that frequent in world languages (Maddieson 1984).

The difficulty of sounds from the intralinguistic perspective can be also established on the basis of aerodynamic explanations. It is possible to measure the amount of effort expended in articulation by counting and comparing the number of articulatory gestures. Boersma (1990) argued convincingly that the concept of

“difficult” and “easy” sounds governs consonants inventories. He accounted for the inventories in simple aerodynamic terms. Articulatory effort is expressed by the presence or absence of gestures. He considered the three candidates, namely *apa*, *ampa* and *api*. *Apa* involves two articulatory gestures: 1) lips close 2) lips open.¹⁹ As a result, there are only two articulatory gestures. *Ampa* involves four gestures: 1) lips open 2) lips close 3) velum closes 4) velic closure removes. *Api* also involves four gestures: 1) lips open 2) lips close 3) aspiration 4) tongue raising. Thus, both *ampa* and *api* feature four articulatory gestures. *Apa* involves only two and wins as the optimal candidate. Therefore, the words with the smallest number of articulatory gestures are preferred. Thus, the hierarchy of articulatory effort can be proposed. Its parameters are voicing, manner of articulation and allophonic processes. He discussed the example of two candidates: *pa* and *ba*, concluding that *pa* is easier to produce than *ba*. *Pa* can be implemented with passive larynx. In order to pronounce the fully voiced *ba*, the speaker has to adjust the width of the glottis to make a closing. The speaker may also lower larynx or slack the supraglottal walls.

Pa involves one articulatory gesture less than *ba* because for *pa* the adjustment of the glottis is not required. In *pa* the increase in oral pressure due to the labial action involves a rise in the intraglottal pressure which pushes the vocal folds apart. Maintenance of the vocal folds vibration during *ba* proceeds under the condition that they are slack. Boersma (1990) made use of the condition which reads: the transglottal pressure must be higher than the one in the glottal area. In the case of the supraglottal closure in *ba*, the transglottal pressure is low. The vocal folds vibrate only if they are slack enough, while the vocal folds tension must be adjusted if the voiced *ba* is produced, whereas no such action is necessary for the voiceless *pa*. In this way, *pa* is easier to produce than *ba* because *pa* requires one articulatory gesture (vocal folds adjustment) less. The issue connected with manner of articulation are also addressed. He claims that stops are easier than fricatives because a ballistic movement is easier than a controlled movement of the articulators. Boersma (1990) quoted the famous metaphor that it is easier to hit a wall than to stop an inch from it (he provides no references). This refers to the different configurations of the vocal tract. Precision is greater for fricatives than stops. In this way, *pa* is preferred over *fa*. This phenomenon can be accounted for by aerodynamics laws. In the case of fricatives, a partial closure is

¹⁹ Perceptually, there are five gestures in every example: 1) labial contour 2) voicing 3) sonorancy 4) continuancy 5) noise.

formed which requires some control from the muscles of the tongue to obstruct the turbulent airflow so that friction results. He considered whether an aspirated stop is easier to produce than an unaspirated one. In production of the aspirated voiceless plosive the vocal folds are abducted nearly to the position required for respiration, namely the /h/ period. Unaspirated /p/ has no glottal gesture and is preferred over *pa*. Finally, a local ranking of gestures is created for selecting the optimal candidate: slack vocal folds, precision and slack vocal folds. *Pa* in terms of slack vocal folds takes no glottal adjustment. *Ba* does as it is voiced. In terms of precision, *pa* wins over *fa* as more precision is needed for fricatives than stops. In terms of spread glottis, aspirated *pa* requires the opening of the vocal folds whereas unaspirated *pa* does not. Therefore, an active glottal opening gesture (spread glottis) is more difficult than the precision needed for a continuant (precision) and that is more difficult than the implementation of obstruent voicing (slack vocal folds).

3.3. The concept of articulatory effort

Articulatory effort belongs to this category of problematic notions in phonology which are, like the syllable, frequently employed and referred to in the literature, but escape a precise definition. Articulatory effort appears to be surrounded by numerous controversies as its very existence is explicitly denied by most phonological theories (especially by the formal ones) and the vast majority of linguists, regardless of affiliation. A general tendency can be observed that the notion of articulatory effort and its phonological implications appear to be grossly underestimated. Moreover, the climate for the notion seems to be highly unfavorable, probably due to the fact that articulatory effort has been understudied. Suffice it to read how reluctantly articulatory effort is recognized in a linguistic dictionary. The attitude towards the notion is reflected by the word choice: “a putative principle of linguistic behavior, by which speakers attempt to minimize the amount of articulatory effort involved in speaking” (Trask 1996: 200). Scarcity of systematized research arises from the generally held preconception that articulatory effort is not subject to measure and, consequently, cannot be quantified at all, let alone expressed by mathematical terms or formulas. Indeed, any references to the degrees of articulatory effort are vague and describe it as *less* or *more* of an articulatory effort. Nevertheless, these controversies fuel numerous

discussions, demonstrating that this notion calls for a more accurate and explicit definition. Since it has been proven beyond doubt that the tendency to reduce effort in all activities performed by human beings is universal and pervasive, a legitimate question in the sense of Zipf (1949 [1972]) can be posed: whether articulatory effort shapes languages as well and what role it plays. Such a query has not been yet addressed in an explicit and exhaustive manner in the phonological literature, but the presence of articulatory effort has been implicitly hinted at.

The first, strict linguistic mention of articulatory effort in the literature was made by Bopp (1889 [1974]) in the 18th century. It has been evoked as an explanatory principle in accounting for language change. Unfortunately, he failed to substantiate his claim in a sufficient way. Articulatory effort as a plausible explanation of explanation of language change has become the prime target of criticisms ever since. After this rather unfortunate debut in linguistics, the notion has been approached with more caution and discussed in the sense of Plato (360 BC [1991]), i.e. as an intrinsic difficulty of sounds (Whitney 1878 [1971]), Baudoin de Courteney 1877 [1972]). Then, the notion has not been taken up for some time, with the notable and isolated exception of Zipf (1935 [1969]), who also investigated the notion in the context of difficult and easy sounds in production. Then, articulatory effort has fallen into oblivion by late 1990s which witnessed the revival of the notion in terms of the H&H theory (Lindblom 1990) and the use of articulatory effort in OT (Boersma 1998). At present, the discussion of articulatory effort has evolved into the direction of articulatory cost and energy (Lindblom 2001, Maddieson 2005). The following subsections review various treatments, theories and models developed in an attempt to explain articulatory effort, listed chronologically. The selection of authors and works perhaps is not exhaustive, but it was determined by their indirect or direct affiliation with Zipf's (1942 [1972]) concept of least effort.

3.3.1. Whitney (1878 [1971])

Numerous linguists link the order of sounds to the greatest or the least difficulty in articulation. This is a simplistic and straightforward interpretation of the easy and difficult concepts in sound production. The practice of sequencing sounds in terms of difficulty in historical perspective can be called into question as the easiest sounds were not

necessarily the first ones (Whitney (1878 [1971])). There are no doubts, however, as to the issue that certain sounds are easier to learn, whereas others are difficult. A scale of difficulty can be set on the basis of observation of what comes first to children, according to Whitney's (1878 [1971]) suggestion: "[s]ome sounds are harder to catch and reproduce than others; and it would be practicable, and highly interesting, to determine by a wide observation and deduction what is the general scale of difficulty of acquisition among alphabetic elements" (Whitney 1878 [1971]: 251). In comparison, for adult speakers there are no difficult or easy sounds per se because they have achieved mastery in their mother tongues.²⁰ It is not intrinsic complexity of sounds: "[t]he difficulty of reproduction lies in the quick and nice transition from one articulatory position of the organs to another" (Whitney 1878 [1971]: 251), but rather the difficulty consists in the challenges posed by rapid speech. Adult speakers experience difficulties caused by time and space limitations in rapid speech: "the succession of different articulating positions, the constant transitions of the organs from one combination to another-these make a modifying influence of higher importance than the differences of intrinsic ease" (Whitney 1878 [1971]: 256-257). It can be speculated that any sounds produced in medial area of the oral cavity are judged to be relatively easier for adults, because they require less effort and the tongue does not have to strain to reach peripheral positions: "[t]he medial sounds, though harder for the untrained speakers to catch and imitate, are found by dexterous speaker a lightening of his task" (Whitney 1878 [1971]: 258). Sounds are more difficult when the distance grows and it has to be covered, like from neutral position of the lips to velum.

3.3.2. Baudouin de Courtenay (1877 [1972])

The magnitude of complexity can be related to general development of language. Baudouin de Courtenay (1877 [1972]) formulated a law which governs the evolution of human speech systems. It "holds that a more difficult sound or group of sounds is replaced in the course of time by one more easily pronounced" (Baudouin de Courtenay 1877 [1972]: 57). Therefore, it can be assumed that easier sounds are preferred to the point of being favored by sound systems. The change is unidirectional, namely

²⁰ Unless these are sounds of a foreign language.

particular languages do not display proclivity for incorporating more difficult sounds in their phonologies.

3.3.3. Zipf (1935 [1969])

The intuitiveness of the notion *articulatory effort* renders it difficult to measure in absolute terms. Phonemes are very well describable in terms of place and manner of articulation, however, quantitative differences are not so easy to capture as the magnitude of complexity of individual phonemes presents a huge challenge. Technically, such a simulation is impossible. There are doubts whether modeling the expenditure of energy involved in production can ever prove viable: “(...) the speech sounds of a given phoneme in a given language differ so widely in expiratory force, duration, pitch, amplitude, and the like in the stream of speech that even a reasonably accurate computation of an average norm seems a practical impossibility” (Zipf 1935 [1965]: 59). Thus, an alternative method of determining the magnitude of complexity can be proposed. It consists in approaching a phoneme as a sequence of gestures: “speech sounds and phonemes may be viewed as constellations, or configurations, or articulatory sub-gestures, arranged partly or completely in sequential order” (Zipf 1935 [1965]: 60). It is noteworthy that already, Zipf 1935 [1965] long ago before Articulatory Phonology writes about gestures. Their number is crucial to establish the amount of effort: the greater the number of gestures required in production, the more complicated the sound is. This method seems to be the only logical option and also allows for comparison of two allophones, such as voicing and aspiration.

The relation between effort and frequency is another logical development of the least effort principle. If a sound is easy to produce, it will be of high frequency of token. More complex sounds will be avoided, i.e. less frequent. Zipf (1935 [1965]) observed this correlation for stops can be observed and accordingly, distributional statements (i.e. whether a given sound is easy or difficult) can be inferred. He noticed that unaspirated stops are more frequent than the aspirated ones. On the basis of this fact, he concluded that unaspirated plosives require less energy. The same holds true for voicing: its lack makes a sound easier. Voiced stops are more difficult, and by the same token, less frequent than their voiceless counterparts. As far as articulatory effort in vowels is concerned, Zipf (1935 [1965]) observed that long vowels have greater magnitude of

complexity than short ones: “*a* represents everything that *a:* represents, *plus* added duration” (Zipf 1935 [1965]: 77), in wake of which, long vowels are less frequent than short ones. The correlation between difficulty and frequency is called *a condition of equilibrium*, where outstanding phonetic difficulty is counteracted by low frequency. On the other hand, high frequency of token equalizes the small expenditure of energy. Zipf (1935 [1965]) suggested that the condition of equilibrium involves the operation of the least effort principle which can be seen in lower tokens for difficult sounds.

A legitimate question can be asked whether he derived frequency from complexity or complexity from frequency and it appears that he suggested that complexity is dependent upon frequency.²¹ The reasoning is simple: people in speech go by meaning and not phonetics of particular sounds. Therefore, it is frequency that triggers reductions. As a consequence, a difficult and less popular sound “changes toward a lesser magnitude of complexity resulting from what may be considered a change toward a greater frequency of token” (Zipf 1935 [1965]: 81). This interpretation “incidentally sheds light on many complex problems in the dynamics of the form and behavior of phonemes in the stream of speech” (Zipf 1935 [1965]: 81). The conclusion that words used frequently will require the least energy seems logical and very much in accordance with the least effort principle. One of the implications of the frequency - complexity effect is the nature of errors. It is obvious that the chances to err are the highest when a difficult or infrequent sound is produced. He used the analogy of the target and failure to approximate the phonemic norm is referred to by him as to *skewness*. In addition to sounds in isolation, Zipf (1935 [1965]) briefly reviews phonotactics, highlighting the effect of one sound on the other in a series: “(...) it is easier to pronounce for example a *d* after *n* than after *m*, yet it is easier to pronounce a *b* after *m* than after *n*. Similarly, *t* is easier to pronounce in the combinations *st* or *ts* than in *tk* or *tm*” (Zipf 1935 [1965]: 96). Complexity magnitude lies in the biological conditioning of the vocal tract. Both /s/ and /t/ are produced in the alveolar region, therefore, in transition from one to the other the configuration of articulators remains unchanged. Thus, homorganic clusters are easier to articulate. Unfortunately, no *Pattern of Clusters* exists as a source of reference. One can only speculate that even in cluster

²¹ This resembles the famous chicken-or-egg dilemma: if a phenomenon is frequent because it is easy or it is easy because it is frequent.

languages, assimilation processes adopt one feature (place or manner of articulation, voicing) to the following or the preceding one.

3.3.4. Lindblom (1990)

Lindblom (1990) developed the hyper- and hypo-speech theory (henceforth H&H), which accounts for phonetic variation. It formalized the interplay between economy in articulation and maximization of acoustic signal's quality. Its basic tenet is that the speaker economizes on articulatory effort when the listener can successfully retrieve acoustic signal with recourse to alternative sources (hypo-speech), whereas expending effort takes place whenever the listener requires maximally specified information.²² The theory stipulates that articulatory effort is expended according to the linguistic situation and claims that distribution of effort varies with respect to circumstances. It predicts that in speech act the listener accommodates the effort expenditure to the listener's discriminating capacities. The adjustment is fine, just sufficient for a successful decoding of the message. The speaker makes allowances for the communicative requirements of the speech situation:

The speaker estimates the running contribution that signal-complementary processes will make during the course of an utterance, and dynamically tunes the production of its elements to the short-term demands for either output-oriented control (hyper-speech) or system-oriented control (hypo-speech). What s/he needs to control is-not that linguistic units are actualized in terms of physical invariants...-but that their signal attributes possess sufficient contrast, that is discriminative power that is sufficient for lexical access (Lindblom 1990: 405).

The theory proposes a continuum of hyper- and hypo-speech and in this way the current communication demand is satisfied. The casual, spontaneous speech is at the hypo end, whereas elaborated variation comes close to the hyper end of the spectrum. This implies that speakers vary the care of articulation across registers. In motherese, foreigner talk and other varieties of emphatic registers, the speaker uses hyper-speech which consists in applying strategies such as repetitions, prolonging duration of utterances and general hyper-articulating of particular sounds. On the other hand, in a more relaxed or familiar

²² By which linguistic knowledge is understood or else the phoneme as a mental image. The listener has representations of certain phrases and the acoustic signal ranks as a support.

environment, the speaker switches to the hypo-speech mode and puts less effort accordingly.

The hypo– and hyper–speech scale constitutes a platform for trade-offs and guarantees that languages effectively balance ease of articulation and clarity of speech (Lindblom 1990). Register as well as situational demands are the governing parameters. As a consequence, the speaker consciously and deliberately controls as well as manipulates the H&H parameter. Thus, the speaker switches between the spectrum ends in order to address the communicative needs. Ideally, the speaker allocates less effort in articulations for those speech chunks which can be easily discriminated by the listener, and, in turn, more effort for more difficult chunks. In poor communication the speaker needs to invest more energy in order to prevent from communication breakdown, whereas in robust communication there is a constraint on effort. The constraint acts against unintelligibility in hypo-speech and comes into force when the listener can discriminate target words or phrases from similar ones. The H&H theory has been extended by Aylett (2000) who demonstrated the relationship between care of articulation and prosodic prominence. Aylett found that the prosodic structures in English also carry the constraint:

Prosodic prominence increases care of articulation and coincides with unpredictable sections of speech [...] leads to a smoother signal redundancy [...] In addition, when variation in prosodic boundaries is controlled for, language redundancy can predict up to 65% of the variance in raw syllabic duration [...] comparable with 64% predicted by prosodic prominence (accent, lexical stress and vowel type) (Aylett 2000).

The H&H theory operates between and within utterances, whereas it seems to fail on the level of a word. The study conducted by Billerey-Mosier (2000) compared the acoustic duration of segments “at and past a carrier’s Word Uniqueness (UP) point” (Billerey-Mosier 2000). The UP is the first or earliest point in a word from which the semantic distinction between that word and the others, i.e. its lexical competitors, can be drawn. Words in the lexicon have in common a string of sounds which happens to be identical (e.g. *street*, *stream* etc.) The point is a string of sounds and is referred to as initial *cohort* (Marslen-Wilson 1987). Beyond the UP, words are identified with a considerable confidence. Thus, any material past the UP point should be redundant, paid less attention to and be subject to systematical phonetic reduction. Billerey-Mosier (2000) examined realization of the *-erry* suffix in words such as *blueberry* whose competitors are *bluebird*, *bluebell* or *blueblood*. In the light of evidence gathered, he

concluded that there are no reductions of articulatory effort behind the UP point. These results demonstrate that articulation, at least on the level of word, does not mirror the listener's processing patterns. If it was so, then the speaker would reduce the *berry* part immediately after its successful recognition by the listener. These findings are supported by means of another study carried out by Sotillo (1997). She implied that the articulation is in general less listener-oriented than it is held by the H&H theory. She investigated the behavior of speakers who were asked the way by tourists and had to pronounce the name of a landmark. The name was known to the speaker, but new to the listeners. Surprisingly, the landmark name was invariably reduced even though there was a new listener who wasn't familiar with the name.

3.3.5. Boersma (1998)

The notion of articulatory effort is conditioned by to the following statement: "we are too lazy to spend any positive amount of effort" (Boersma 1998: 149). A constraint follows that articulations which require less effort are favored and this is supposed to be a universal constraint. Boersma (1998) arrived at his parameters by means of criticizing the previous treatment of effort which operated with one variable. For him, articulatory effort is not a monolith unit but rather is composed of several articulatory parameters. The question arises whether articulatory effort can be measured at all and which measures should be adopted. The measures themselves derive from physics, in particular the use low-level kinematic parameters such as duration, mass, energy, distance, precision, coordination, speed etc. The parameters are selected from the whole range. For Lindblom (2001), for instance, effort should be measured by means of the jaws' mass, damping and velocity, whereas for Boersma, articulatory effort "depends on at least six primitives: energy, the presence of articulatory gestures, synchronization of gestures, precision, systemic effort, and coordination" (Boersma 1998: 149) which are ranked both globally and locally by languages. Energy is defined as the total work done by muscles in covering distance from neutral to target position. The former position means that the tongue passes from posterior to anterior regions of mouth and in passing it reaches the zero point. Naturally, energy is also connected with duration: the longer a sound, the more energy is needed. The presence of articulatory gesture serves as "a first rough measure of the organizational effort of an utterance" (Boersma 1998: 151).

Indeed, one can count the gestures required for the articulation of /t/ vs. /d/ and conclude, that all else being equal, /d/ involves one gesture more: voicing. Synchronization is the timing of two articulatory contours. The effort saving imperative implies that “two articulatory contours on different gestural tiers like to be far apart” (Boersma 1998: 154). Precision relates to position and environment and is often traded for minimal articulatory effort: “[i]n certain environment, a certain articulator does not work up the precision to put itself in a certain position” (Boersma 1998: 156). Stops require less precision than a fricative. This is corroborated by statistics: in the languages of the world, stops are preferred over spirants (Maddieson 1984). Systemic effort consists in maximum using of features at the speaker’s disposal. Boersma (1998) understands these features as tricks in articulation, i.e. specific way of articulating particular sounds of a language system. Coordination is defined as “combination of articulatory tricks to be learned” (Boersma 1998: 156). These two last parameters of systemic effort and coordination merge into one. All the proposed parameters exhaust the list of articulatory effort factors.

3.3.6. Maddieson (p.c.)

The problems connected with effort measure stem from the fact that the assessment of effort operates on an intuitive level. Consequently, effort parameters are assessed intuitively. Therefore, an intuitive proposal can be put forward in terms of effort parameters: distance, coordination and precision (Ian Maddieson, p.c.).²³ The latter one appears to have a strong motivation in articulation that is the degrees of air obstruction in the vocal tract. It is felt that the sounds which require high precision would be judged effort-costly. The famous stop–fricative metaphor: “it is easier to run into a wall than to halt an inch in front of it” (source unknown, in Boersma 1990) involves that it is easier to form a constriction, to build up the air pressure and then to release it explosively than to control and adjust the partial closure which should result in friction. The distance between articulators plays a role since places of articulation are located within all of the

²³At the beginning of his professional career, Maddieson took an interest in articulatory effort. In particular, he wanted to research effort in child language for which he developed a measure. He applied for a grant for the articulatory effort project but didn’t receive it. He succeeded, however, in obtaining a grant for his *Patterns of sounds* (1984) and abandoned the research on articulatory effort (Ian Maddieson, p.c.). Nevertheless, he shared some ideas about effort, contributing substantially to the present discussion.

vocal tract areas, ranging from lower and upper lip to glottis. Thus, certain sounds are located on the peripheries of the central point of the oral cavity which relatively increases the difficulty. Articulation of a front low vowel preceded or followed by a bilabial sound appears to involve a shorter distance than that of velar stop and a low front vowel. By the same token, this parameter favors the schwa sound as there is no distance to cover. Coordination increases with the number of articulatory gestures involved, a substantial number of gestures raises production difficulty as more actions must be coordinated. Maddieson (2005) discussed the case of displaced articulations which are considered complex as they involve the movement of an active articulator in a direction other than the passive articulator opposite which it lies at rest. For instance, in retroflex the tongue reaches the closure at the position further up than at rest, dentals reach the closure at the position more to the front than at rest. Consequently, labiodentals are the example of a displaced articulation. If one adopts the frequency explanation, labiodental sounds are more difficult, so they should be rare, whereas bilabial should be frequent as they do not involve displaced articulation. On the other hand, labiodentals involve the movement of only one lip, whereas the production of bilabials involves moving both upper and lower lip. The data obtained with the help of EMMA on timing are crucial: In the Ewe language the movement of the upper lip is subtler.²⁴ Timing of rise of the upper lip during constriction does not suggest that this is occurring merely because the lower lip pushes the upper backwards after the articulation (Löfqvist – Gracco 1996). Elevation of the lower lip occurs prior to constriction, first the lower lip lowers, whereas the upper continues to rise. This move ensures that a fricative is produced and not a stop, while timing reversal could produce the unintended result. Therefore, active adjustment of crucial closure takes place which suggests that perhaps in terms of coordination bilabials are more complex than labiodentals.

A caution must be taken in assessing effort according to only one parameter since the assessment may be altered in connection with the other ones, namely precision and distance. The parameters considered in isolation may produce a different result

²⁴ EMMA, also referred to as or EMA, or EM(M)A stands for Electromagnetic Midsagittal Articulography. It is a device which gains an insight into the oral cavity. It is the most modern device of that type, in comparison with line X-ray, X-ray microbeam, electropalatography (EPG), Magnetic resonance Imaging (MRI) and ultrasounds. EMMA generates data by means of transmitter coils (responsible for magnetic fields at different frequencies), transducer coils (convert the magnetic fields into currents) and current (gives signal which is fed through a special interface data into a PC in the digital form).

when in combination with the remaining ones. It would also be interesting to explore whether these parameters are of equal weight or are arranged hierarchically, for instance, distance being the least important. Ideally, one should investigate all possible low-level kinematic parameters and seek correlation between them, e.g. if parameter A increases, then parameter B decreases, whereas parameter C increases as well. Moreover, the question arises whether coordination, precision and distance exhaust the list of possible parameters, as speed of articulation, to mention one example, could be also taken into account. The number and hierarchy of effort parameters remains still an open question, partly due to the fact that effort is an intuitive correlate and partly due to lack of proper measurement tools which could verify each of these effort parameters. Some methods of measuring articulatory effort, however, have been pioneered. Nadler et al. (1987) measured articulators displacement via microbeam tracking; currently, EMMA appears to be a promising tool in effort measurements. Lindblom — Moon (2001) developed a metrics of heat energy (measured in Jules) generated in speech so that judgments of complexity are independent from frequency.

3.4. Articulatory gestures

The notion of articulatory gestures derives from the motor theory of speech perception (Lieberman — Mattingly 1985). The very idea of articulatory effort necessitates certain references to articulatory phonology so that gestures can be incorporated on aerodynamics grounds into the concept of least effort. The concept of articulatory gestures was first explicitly postulated by Liberman (1957) and Liberman and Mattingly (1985), although earlier mentions of gestures can be found in Zipf (1935 [1965]). They hypothesized that these articulatory movements are the platform of both production as the speaker produces a sequence of movements and perception since the hearer extracts the acoustic signal from them: “[i]n its extreme and old-fashioned form, this view says that we overtly mimic the incoming speech sounds and then respond to the proprioceptive and tactile stimuli that are produced by our own articulatory movements” (Lieberman 1957: 122). They also claimed that the speaker produces a series of articulatory movements which overlap temporally. These movements are subject to the perception of speech as “the intended, phonetic gestures of the speaker,

represented in the brain as invariant motor commands” (Lieberman – Mattingly 1985: 2). In similar vein, the Articulatory Phonology model as developed by Browman and Goldstein (1986, 1992) emphasized the motor aspect of phonological organization. They undertook the task of searching for phonological primes and formalized the model of Articulatory Phonology, challenging the major assumptions of the segmental and linear phonology models. The segmental one operated with the category of segment as the smallest unit in phonology, whereas the linear one hypothesized that segments are arranged into a sequence and are the bundles of features (Chomsky – Halle 1968). These two models failed to account for a variety of phonological facts such as segments overlapping and temporal organization of speech (Browman – Goldstein 1986, 1992). Therefore, the assumption that features are the basic units of speech is challenged. They claim that features are described in phonetic-impressionistic terms: for instance, the feature *strident* is an acoustic one (Chomsky – Halle 1968). The advent of new technologies such as pellets and microbeams which track articulatory movements provided the opportunity of a more precise, physical and non-impressionistic or acoustic measurements (Fujimura 1981, Kiritani 1986). Features are static and thus fail to focus on the dynamic aspect of speech.²⁵ In order to remedy this problem, Browman and Goldstein (1986, 1992) adopted the elements of the task dynamic model (Kelso 1984, Saltzman – Kelso 1987).

Articulatory Phonology formalized the notion of articulatory gestures. They are the formations of constriction in the vocal tract in space over time and can be defined as “characteristic patterns of movement of vocal tract articulators, or articulatory systems” (Browman – Goldstein 1986: 223). Trask (1996) defines gestures as follows:

In some approaches to phonology, any of the several partly independent components into which the articulation of a segment may be decomposed. One simple approach is to decompose each segment into a laryngeal gesture, including everything going on in the larynx, and a supralaryngeal gesture, including everything going on in the mouth and the nasal cavity. Most developments of the idea recognize more than two gestures, however, usually including some other gestures which are subgestures, of other superordinate gestures. One typical proposal is to divide a segment into a *categorical gesture*, consisting of *phonatory* and *initiator subgestures*, and an *articulatory gesture*, consisting of *locational* and *oro-nasal subgestures*; each of the four subgestures is the domain of certain distinctive features or components (Trask 1996: 156-157).

²⁵ Even the IPA names the sounds of the world’s languages after the gestures: labiodental fricative, not features: strident-anterior-continuant.

Furthermore, gestures simplify the account of phonological facts: “gestures are autonomous structures that can generate articulatory trajectories in space and time without any additional interpretation or implementation of rules” (Browman – Goldstein 1986: 223). Within the framework of Articulatory Phonology, gestures have the status of phonological primes which temporally characterize articulation movements on abstract level. Therefore, a bilabial closure gesture implies a pattern of movement which repeats itself and is linguistically significant. Gestures viewed as phonological primes are capable of introducing the lexical contrast: *add* vs. *had*, *bad* vs. *pad*. However, there is no correspondence between segments and gestures. One segment /m/ may involve multiple gestures: velic and oral. Gestures also do not have a one-to-one correspondence to features, but rather, “they represent organized patterns of movements within oral, laryngeal and nasal articulatory systems” (Browman – Goldstein 1986: 225). This marks the evolution from phonetic-impressionistic descriptions, e.g. *tongue rising* to articulatory movements over time.

Articulatory gestures are specified by the vocal tract variables as there is a set of parameters which organize the constrictors’ movements in space. The variables coordinate the constriction in terms of the constriction’s location (abbreviated to as CL) and the constriction degree (CD). The latter variable roughly corresponds to the manner of articulation, whereas the former to the place of articulation. There is also a set of parameters which organize the constrictor’s movements over time. This particular vocal tract variable is called *stiffness*. It specifies the time which a gesture needs to reach the target in space. Thus, the spatial and temporal variables are related to each other. The vocal tract variables are organized along the lines of the major articulators which form the constriction. These constrictors are velic, nasal and laryngeal, while only the oral constrictor is specified by CL (Constriction Location) because the larynx and the velum are constrictors per se and do not need any further specification as to their localization. In turn, the oral constrictor is divided into two major active articulators, namely the lips and the tongue. The tongue is further subdivided into the tip, the body and the root. These division lines are paired with the CD (Constriction Degree) variables since the tip of the tongue may assume the apical or laminal shapes. The CD variable allows for introducing the lexical contrast. Values such as [closed], [critical], [mid], [narrow] and [wide] bring the contrast in meaning of lexical items. The contrastive function consists in the presence or absence of a gesture. However, the above values do not correspond to the traditional, phonological features of Chomsky and Halle (1968). Features do not overlap since they

are separated values. Gestures in Articulatory phonology are embedded within the task dynamics of Saltzman and Kelso (1987) which entails that the gestures' overlap over time blends their influences on common articulator. Thus, gestures capture an array of phonological data. Under the gesture approach "utterances are modeled as organized patterns (or constellations) of gestures, in which gestural units may overlap in time" (Browman – Goldstein 1992:155). These constellations have explanatory power in phonological variations such as allophony, connected speech or speech errors. Allophonic variation results from the timing relations between particular gestures (Gafos 1999):

[...] complete overlap between a velic and an oral gesture results in the percept of a single consonant [n]. When the velic gesture is slightly slid leftward (i.e. anticipated), the resulting overlap is partial and gives the percept of the so-called 'pre-nasal stop' [ⁿd] [...] a minimal overlap would give rise to the percept of a velic gesture dissociated from the oral gesture, the cluster [nd] (Gafos 1999).

Chapter Four

Effort management: the experiments

4.1. The aim of the chapter

The aim of the chapter is to provide evidence for the principle of least effort in articulation on the basis of experiments. First, the chapter proposes to substitute the term *least effort* with *effort management*. It also presents the empirical part of the thesis which includes three experiments. Section 4.2. describes the experiments in terms of their predictions and methodology. In particular, it discusses the adopted measure of effort established by means of the effort parameters as well as their status. Then, the chapter outlines the statistical aspect of the experiment. Section 4.3. clarifies the statistical methods used in the experiments, such as the sample size and the confidence interval.²⁶ Next, the chapter reports on the results of the experiments. Section 4.4. presents the results in a graphical way and discusses their relation to the tested predictions. Finally, the chapter discusses the significance of the results and proposes the attention hypothesis. Section 4.5. concludes that effort management plays a predominant role in the speaker-listener communication act. The conclusion is additionally supported by the three studies in text messages demonstrating that in texting the effort of written communication is in fact deliberately managed.

²⁶ All the statistical analysis in the thesis is performed with substantial help of Hubert Pikosz (Hubert Pikosz, p.c.) from Poznan University of Technology.

4.2. The principle of least effort as effort management

4.2.1. Explanation of effort management

The principle of least effort in the understanding of Zipf (1949/1972) is the central point of interest of the thesis. He argued that the principle governs the totality of human behavior. Since speech constitutes a distinguished manifestation of human behavior, it also is subject to the principle of least effort. In view of the fact that least effort originates from least work, articulation is the work to be performed in the case of speech. Thus, the main interest of this chapter is to pursue the idea of effort in relation to articulation. Articulation is used here in the sense of Browman and Goldstein (1986, 1992) and Bussmann (1996): “intentional movement of the primary articulators for the creation of speech sounds, including those organs involved in the airstream mechanism and phonation” (Bussmann 1996: 35). The intentionality of movements means that the speaker is capable of controlling deliberately and tuning the amount of effort. The very name of the principle explored by the present thesis, however, calls for a necessary terminological revision. Therefore, the thesis proposes to use the term *effort management* instead of *least effort*. This proposal has four reasons. Firstly, the original name of the principle appears to be selected in an infelicitous way. The term *least effort* seems inadequate or, indeed, not very precise. In fact, its vagueness makes it difficult to understand the principle and its implications. Specifically, the notion of least effort leads up the garden path, erroneously inviting the interpretation of the simplest or the shortest way of tackling any task. This interpretation, however, runs counter to Zipf’s (1949 [1972]) intentions. Thus, the current term *least effort* fails to do justice to the meaning it denotes. Secondly, the present hypothesis stipulates that in the light of the Zipf’s (1949 [1972]) principle articulatory effort is managed. This stipulation means that effort is expended in an effective way, with a view of lessening the workload in the long run. The term *least effort* suggests that the workload is rather avoided, unlike the term *effort management* which stands for an optimal expenditure of effort and a deliberate selection of a path to follow. *Effort management* implies a cyclical process with monitoring of effectiveness. Thirdly, effort management derives from the communicative function of language. It observes the imperative that a message must be conveyed as quickly as possible (Sweet 1891 [1960] and Passy 1890). Therefore, the speaker manages effort in order to communicate effectively at the relatively lowest articulatory cost. In the event when the speaker fails to include a sufficient number of

signals to be decoded by the listener, the effort is mismanaged. The speaker is usually requested to repeat the part where he expended too little effort and repetitions double or multiply effort.

4.2.2. The predictions of the experiments

The aim of the thesis is to explore the relation between the speaker and the listener in terms of articulatory effort which is associated with the speaker. Naturally, the listener also makes effort when decoding and interpreting the message. Speech, unlike perception, is an external process and may be assessed objectively. In the case of perception, one would need to rely on the listeners' judgments which is a more subjective method. Thus, the access to effort made in speaking motivates the choice of the focus on articulatory effort. Articulatory effort is considered to be the negotiating platform between the speaker and the listener on which the two conflicting interests must surface. It is also speculated that the negotiating platform does not operate in an unconstrained fashion, but is subject to certain general principles. In the light of the Zipf's (1949 [1972]) principle, the speaker should be making the least effort, whereas in the light of the functionalist principle (Sweet 1891 [1960] and Passy 1890), the speaker manages it not according to the needs of the listener but to his own in order to be understood. The need of the speaker to be understood also derives from a function from speech act theory called *securing of uptake* (Austin 1962). This concept describes an illocutionary act in which the speaker must make it clear to the listener that the act is performed (Austin 1962). Following the three principles, the two predictions are formulated: the first prediction is that the speaker makes little articulatory effort when s/he assumes that the listener does not need the full range of acoustic signals. The low degree of articulatory effort is referred to as hypoarticulation (after Lindblom 1990). This happens in the situation when the speaker articulates a word of high frequency, which s/he thinks is known to the listener. On the other hand, the speaker increases the effort if s/he fears that the listener will force him/her to repeat the continuum of signals (the second prediction). The high degree of articulatory effort is referred to as hyperarticulation (after Lindblom 1990). This happens when the speaker articulates a word of low frequency which s/he assumes is not known to the listener. Repetition is in principle more costly than hyperarticulating. Therefore, the speaker hyperarticulates

only in order not to repeat and not because s/he takes into account the needs of the listener. These two predictions lead to a conclusion that indeed the speaker manages the effort in the sense of Zipf (1949 [1972]). To use the Zipf's example, when the speaker assumes that the listener would not understand, s/he builds the tunnel through the mountains (by hyperarticulation) because s/he saves the energy s/he would expend if s/he climbed the mountains (by hypoarticulation). These predictions derive from the very function of language and the claim that the natural tendency of human beings is to minimize effort in every preformed activity, speech included. They both seem to be highly plausible, reasonable and their universal nature had been proven (cf. Chapter One). Nevertheless, the predictions made on the basis of the principles should be verified empirically.²⁷ Therefore, three experiments were designed and conducted in order to provide empirical support to the two predictions.

4.2.3. Methodology of the experiments

As far as methodology is concerned, repetitions and corrections constitute a manageable research method since acoustic measures can be compared and an effort index can be established: vowel quality in terms of centralization as well as tongue displacement for consonants (Moon – Lindblom 1994). In other experiments, an increase or decrease of effort can be predicted by means of other method: whether the speaker speaks to a native or a non - native speaker (Ian Maddieson p.c.). Effort is very well manageable in high frequency words as less effort is invested and they are subject to reduction. The method of researching low frequency words against these of high frequency does not appear to be effective in testing effort management due to the fact that high frequency words by definition display a greater reduction than those of lower frequency (Bybee 2001). Rather, it was decided to use the simple method of repetitions as a way of testing the above predictions. In a text, the word was repeated so that its tokens were embedded. The tokens of one word ensure the possibility of observing whether articulatory effort changes or remains on a stable level throughout all the tokens. The method of observing the behavior of articulatory effort is to manipulate the semantic context in which the tokens appear. Therefore, the tokens which were used in the same

²⁷ There is a conviction, however, that if one selects the right principles, the claims must be true and need not to be verified at all (Spinoza 1662, 1667 [1955]).

context were expected to be hypoarticulated. According to the first prediction, if the speaker uses a word with the same meaning a few times in a conversation, the speaker thinks that the listener already knows the word and makes no more articulatory effort than absolutely necessary. When the semantic context changes, the listener may be at a loss about the subject of the conversation. It is speculated that the speaker signals the change of the context and hyperarticulates the word, as the second prediction predicts. This signaling by means of hyperarticulation, however, is not altruistic at all or dictated by the care for the listener. The speaker makes more articulatory effort consciously because s/he wants to be understood at all times, especially when the context changes. Otherwise, the listener is confused and likely to make sure if s/he follows the conversation by asking the speaker. This uncertainty would result in repetition on the part of the speaker and multiplies the effort.

Therefore, for the sake of the first experiment, a text was designed in which the four tokens of one word were used in two different contexts. The first three tokens were used in one context (context A), whereas the fourth and final token was used in a different context (context B) and received the meaning different from the one introduced originally. In order to employ the repetition method, the text had to satisfy a number of criteria. As far as the repeated word is concerned, it had to be a homonym (since semantic context is stipulated to govern articulatory effort), but representing the same grammatical category (pairs such as *shed* as N and V were excluded). The word *web* means *the internet* in the context A (the first three tokens), whereas in the context B it denotes *the spider's trap* (the final token). As far as the phonetic environment is concerned, it must have been controlled and the same in each token so that only the semantic context and not the phonetic one could motivate the changes in articulatory effort. The text of the first experiment, employing the repetition method, is as follows:

The world wide web offers numerous resources. I use the world wide web in my work. I used the world wide web in school assignments on the wide web of deceit surrounding the Watergate scandal and on the gap in earnings of males and females.

Consequently, the repetition method used in the first experiment would verify these two predictions:

Prediction 1: the articulatory effort should decrease along with the tokens of a word used in the same context. In the first experiment, the subsequent tokens of the word in question should be hypoarticulated according to the least effort principle.

Prediction 2: the articulatory effort should increase in the tokens of a word which appear in a different context. In the first experiment, the final token of the word in question used in a different context should be hyperarticulated because repetition is in principle more costly than hyperarticulating.

4.2.4. The measure of articulatory effort

Articulatory effort can be measured in various ways. One of them consists in tracking the movements of articulators. This way combines the parameters traditionally ascribed to biomechanical effort, i.e. precision, distance etc. However, articulator tracking is a relatively new development and the methodology of research is not yet well established. Besides, the results obtained with the help of EM(M)A raise certain objections. Maddieson (2005) employed the notion of *displaced articulation* (cf. 3.2.6.), arguing on the basis of the data obtained by means of EM(M)A that timing of the lip movement is crucial in production of bilabials. While bilabials explained by means of the order of lip movement demonstrate the primacy of the upper lip, it turns out that the same may be argued for the lower one. Therefore, an explanation which is applicable to two opposite hypotheses does not in fact explain anything. This may suggest that the EM(M)A data concerning the articulator movements is far from conclusive. Thus, it was decided to depart from the tracking of articulators approach and to pursue an alternative measure of articulatory effort. The measure of articulatory effort is referred as to *effort measure*, whereas the parameters by which articulatory effort can be measured are referred to as *effort parameters*. Maddieson (p.c.) has suggested vowel duration as a potential effort parameter and a manageable way to measure articulatory effort. It is felt, however, that there are more parameters which contribute to an increase or decrease in articulatory effort. Consequently, the three remaining parameters were added to the parameter 1 (vowel duration): vowel centralization in terms of parameter 2 (vowel height) and parameter 3 (vowel position), final obstruent devoicing in terms of parameter 4 (closure duration and its voicing) and the degree of spirantization (parameter 5). The first three parameters concern vowels, whereas the fourth and the fifth ones concern consonants so that a balance is maintained. The selection of parameters was partly intuitive, partly determined by phonetic facts. The parameters are calculated on the basis the data which

can be derived from a spectrogram by means of Praat, the package software (Boersma – Weenink 2005). Two more parameters could be added: the intensity of the sound and voice quality. However, they were rejected on the following grounds: sound intensity can increase just because the subject has accidentally moved his/her head; the background noise has amplified the signal or the speaker kept an inappropriate distance from the microphone. Voice quality might potentially serve as a measure of effort but it was also rejected as the changes in voice quality (F4) might be accidental and induced by factors independent from the text of the first experiment. The subjects might be tired or irritated by the number of readings or the length of the text.

4.2.5. Discussion of effort parameters

Parameter 1 (vowel duration): it is stipulated that there is a relation between vowel duration and articulatory effort: the longer the duration of a vowel, the more articulatory effort was put in its production. The durations of four vowels in the subsequent word tokens were compared. There are three clues which help to establish the duration of a vowel: the darkening of formants related to acoustic energy (marks the finishing point of the vowel but it does not concern all formants to an equal degree due to the fact that F3 loses its energy while F2 and F3 are still well visible), the waveform amplitude (it is high and regular for vowels) and voicing (marks the vowel starting point).²⁸

Parameter 2 (vowel height) and parameter 3 (vowel position): the central position assumed by the tongue is the most natural and effortless. The higher the degree of vowel centralization, the less effort was invested. Since no unit of centralization was established, it is proposed that the formant frequencies should be compared. F1 determines the tongue height, whereas F2 determines the tongue position. The values of the frequencies obtained for vowels in subsequent tokens were compared. The differences in formant values showed the degree of vowel centralization.

Parameter 4 (obstruent devoicing): devoicing demonstrates an increase in effort, whereas voicing shows a decrease in effort. This is assumed in the case where all the sounds in immediate vicinity are voiced as there is no need to readjust vocal folds for the wide configuration. Consequently, in the environment of voiceless sounds voicing

²⁸ The clues were suggested by Geoffrey Schwartz (p.c.) who provided instructions how to read correctly the duration of a vowel.

would involve a decrease in effort. The former case was considered for the sake of the devoicing parameter and researched in the text of the first and second experiments. Traditionally, three degrees of obstruent realization are identified: fully voiced, partially voiced (devoiced) and voiceless (or two-just voiceless and voiced). There is an ongoing debate, however, on the concept of (in)complete neutralization of contrast. Traditionally, the three phonetic categories are: fully voiced, devoiced and voiceless. The difference between the devoiced and voiceless category is indisputable. However, the treatment of the difference depends on the theoretical bias. Some scholars disregard the difference equating voiceless with devoiced, some build on that distinction. Experimental studies focus on slight perceptual differences between the devoiced and voiceless categories. In Polish, there is an orthography bias and the perception of *kod* vs. *kot* may be influenced by the spelling. Therefore, a study on Catalan which is free from the orthography bias was undertaken (Slowiaczek – Dinnsen 1985). It turned out that the speakers could not hear the difference between the devoiced and voiceless categories. In the thesis, the claim that devoiced and voiceless categories are the same is adopted due to the body of evidence (Manaster Ramer 1996, Port – O'Dell 1985). Final obstruent devoicing consists of two parameters: closure duration (the duration of the closure/silence period lasts longer in the case of voiceless sounds and shorter for their voiced counterparts) and closure voicing. Closure duration as a measure of devoicing is a well-established phonetic fact. Lisker (1957) observed that “[c]losure durations for *p* fall in the 90-140 msec. range, with an average value of about 120 msec., while values for *b* vary from 65 to 90 msec., with an average of 75 msec.” (Lisker 1957: 47). In similar vein, Dalcher (2006) makes use of voicing and release burst as acoustic correlates of lenition in her study of Florentine Italian. As far as closure voicing is concerned, voicing is a dynamic process, the airstream has not been cut off completely and during the closure period still the F2 and F2 are visible. In Praat, voicing is shown as pulses (the blue lines). For voiced sounds, the closure voicing lasts longer. Moreover, the amplitude for voiceless sounds goes down because the closure is quiet.

Parameter 5 (spirantization degree): the measure whether an obstruent in question was realized as a stop or as a spirant. Stops require more precision in production in comparison to fricatives, thus if the stop is spirantized, it means it was hypoarticulated and required less articulatory effort in terms of precision. LaVoie (2001) conducted studies of intervocalic voicing of voiceless stops and intervocalic spirantization of stops in English and Spanish which “show that [...] the phonetic

parallels to phonological voicing and fricativization are short duration and incomplete stop closures. Decreased closure duration of intervocalic stops gives rise to a vocalic percept in the absence of vocal-fold vibration, while stops with incomplete closure can be perceived as fricatives” (Blevins 2004: 147).

4.2.6. The status of the parameters

Under the thesis’ approach, the effort measure is a total of five parameters. It is assumed that all five parameters contribute to overall effort. It does not mean, however, that effort measure is merely a sum of the parameters. It is impossible to add duration and frequencies since they constitute different values. In similar vein, a linear correlation between all parameters (e.g. by means of the Pearson significance correlation coefficient) cannot be investigated due to the fact that an increase in vowel duration (parameter 2) has no effect on its centralization (parameters 2 and 3). Likewise, the degree of spirantization (parameter 5) does not affect obstruent devoicing (parameter 4). These parameters describe independent articulatory gestures and as such are not comparable as far as the relation between them is concerned. Instead of a correlation, it would be relevant to establish a relative contribution of a given parameter to the overall effort measure. It would answer the question if the parameters are equally representative for the effort parameter i.e. if they show in an equal way the amount of articulatory effort. If they demonstrate a similar tendency, they all equally contribute to the overall effort. If they differ, they will be assigned a greater/lesser role in articulatory effort, depending on the nature of the difference.

4.3. The statistical aspect of the experiments

4.3.1. The method of establishing the sample size

The results obtained from a single recording may be a matter of accident and are not reliable. In order to establish the sample size which allows obtaining reliable results, eleven readings were made by a native speaker of English. The number of readings is

referred to as sample size. Next, the modified coefficient of variation was calculated on the basis of the sample size. (Oktaba 1966). It was calculated for the mean value of the first parameter (vowel duration) results obtained from eleven readings for increasing number of measurements (from 2 to 11). The coefficient of variation is the relation of standard deviation of the particular result to the mean value, given in per cent (Oktaba 1966). The coefficient is described by the following formula:

$$W = \frac{\sigma}{\bar{x}} \cdot 100\%$$

W- the coefficient of variation

σ - standard deviation

\bar{x} -mean value

The modified coefficient of variation (the coefficient of variation for the mean value) for an analyzed parameter (here the vowel duration parameter) is the relation of standard deviation of the mean value to the mean value, given in per cent (Oktaba 1966). The modified coefficient is described by the following formula:

$$W^* = \frac{\sigma^*}{\bar{x}}$$

W* - the modified coefficient of variation

where

$$\sigma^* = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$$

so

$$W^* = \frac{W}{\sqrt{n}}$$

n-the sample size

It is expected that an increase in the sample size decreases the value of the modified coefficient of variation. This does not hold true only if there is a substantial difference between the results of subsequent measurements. Beyond a certain number of measurements the decrease in the modified coefficient of variation is insignificant. Therefore, it is assumed that further increase in measurements does not affect the reliability of result. In other words, the modified coefficient of variation establishes the sample size which produces consistent results.

Table 1. Modified coefficient of variation (speaker 1).

sample size	token 1	mean duration	standard deviation	W*
1	0,101544			
2	0,119155	0,110350	0,012452858	7,98%
3	0,092880	0,104526	0,013388975	7,40%
4	0,099107	0,103172	0,011262862	5,46%
5	0,096485	0,101834	0,010202005	4,48%
6	0,094489	0,100610	0,009605036	3,90%
7	0,096617	0,100040	0,008897096	3,36%
8	0,098333	0,099826	0,008259173	2,93%
9	0,083508	0,098013	0,009448521	3,21%
10	0,087604	0,096972	0,00949685	3,10%
11	0,103710	0,097585	0,009235703	2,85%

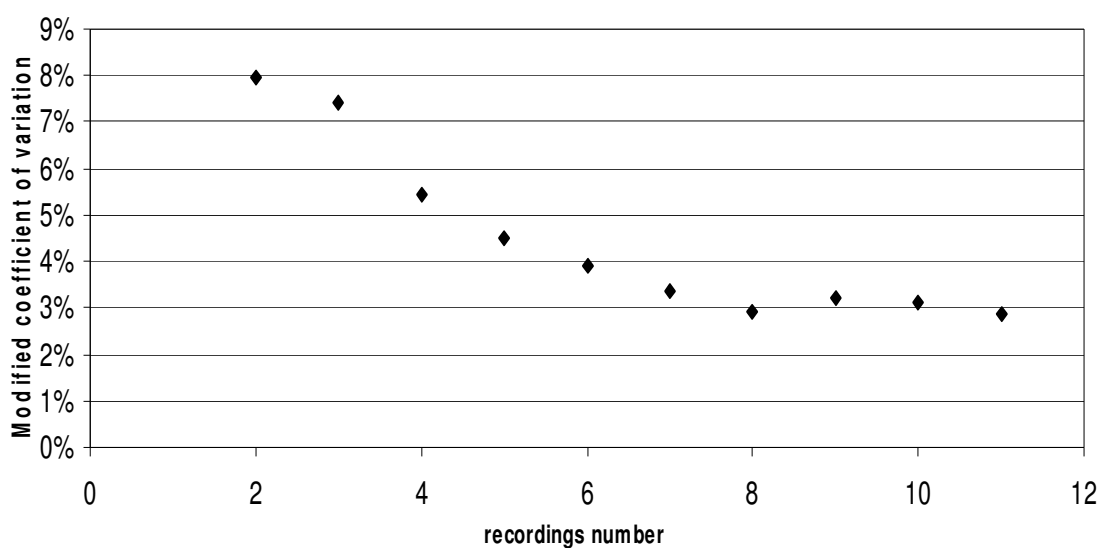


Figure 4. Modified coefficient of variation (speaker 1).

Figure 4 illustrates the modified coefficient of variation for the first speaker. It indicates that the desired sample size is 7. This means that beyond the seventh reading the results become consistent. This is why the speakers are asked to read the text seven times since any number below seven would produce unreliable results. For the sake of the sample size, the text of the first experiment containing four tokens of one word was read eleven

times. Next, the durations of the vowels in the tokens were compared and calculated with respect to the modified coefficient of variation. The results for tokens 1, 3 and 4 were the same, only the token 2 indicated three readings as the sample size. It is recommended in the literature to adopt the highest sample size. Thus, the modified coefficient of variation established the sample size on the level of seven readings. For comparison, the modified coefficient of variation was calculated for another speaker:

Table 2. Modified coefficient of variation (speaker 2).

sample size	token 1	mean duration	standard deviation	W*
1	73			
2	56	64,500000	12,02081528	13,18%
3	53	60,666667	10,78579312	10,26%
4	65	61,750000	9,069178574	7,34%
5	44	58,200000	11,16691542	8,58%
6	59	58,333333	9,99333111	6,99%
7	50	57,142857	9,651054717	6,38%
8	42	55,250000	10,41633333	6,67%
9	51	54,777778	9,846036992	5,99%
10	53	54,600000	9,299940263	5,39%
11	45	53,727273	9,285374619	5,21%

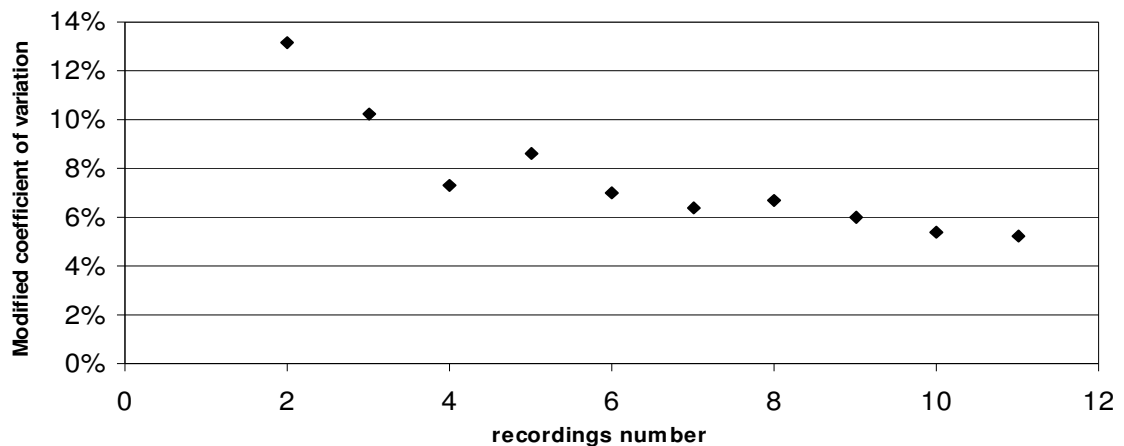


Figure 5. Modified coefficient of variation (speaker 2).

Figure 5 illustrates the modified coefficient of variation. It indicates that the desired sample size is 6. Nevertheless, the number of recordings was adopted for the speaker 1 and all other speakers recorded the text seven times.

4.3.2. Confidence interval

When the results are obtained, the problem of measuring error (or, the problem of chance variation) must be eliminated. The confidence interval determines the percentage of measuring error (Volk 1965). When the intervals overlap, the results are within measuring error and are not reliable. Therefore, the confidence interval was calculated. The standard deviation is multiplied by the t-Student coefficient appropriate for the sample size (Bronstein – Semendjajew 1996). Confidence interval is described by the following formulas (Volk 1965):

For the results:

$$\frac{1}{2} L_{09} = t_{09}^{n-1} * S$$

For mean results:

$$\frac{1}{2} L_{09} = t_{09}^{n-1} * \frac{S}{\sqrt{n}}$$

L-confidence interval

S-standard deviation

L_{09} -confidence interval at the significance level of 0,1

t- t-Student coefficient

n-1 – for less than 1 degree of freedom

n-sample size

Table 3. Quantiles $t(p, \nu)$ of p order of Student's distribution with ν freedom degree.

	p				
ν	0,80	0,90	0,95	0,98	0,99
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,859	2,306	2,897	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169

11	1,363	1,795	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,054
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,724	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,055	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,312	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750

For the sample size of seven recordings the t-Student coefficient is 1,943 (n-1). The confidence interval is calculated at the significance level of 0,1. This means that the measuring error is 10 per cent. Since the significance level is 0,1 for the calculations, the degree of freedom is 0,90. In fact, the half of confidence interval is calculated.

4.4. The empirical evidence

In order to verify the operation of effort management, three experiments were conducted. The results they provided are described in the sections below accordingly.

4.4.1. The results of the first experiment ²⁹

The first experiment was developed in order to test the two predictions formulated in section 4.2. on the basis of least effort and language's function. The text of the first experiment is as follows:

The world wide web offers numerous resources. I use the world wide web in my work. I used the world wide web in school assignments on the wide web of deceit surrounding the Watergate scandal and on the gap in earnings of males and females.

The two predictions may be now formulated in relation to the specific text and the specific words.

Prediction 1: the articulatory effort should decrease along with the subsequent tokens of the word *web* used in the same context. Therefore, the subsequent tokens of the word in question should be hypoarticulated because semantic context governs articulatory effort according to the least effort principle.

Prediction 2: the articulatory effort should increase in the final token of the word in question which appears in a different context. Therefore, the final token of the word in question used in a different context should be hyperarticulated because repetition is in principle more costly than hyperarticulating.

First, the parameter 1 concerning vowel duration is discussed.

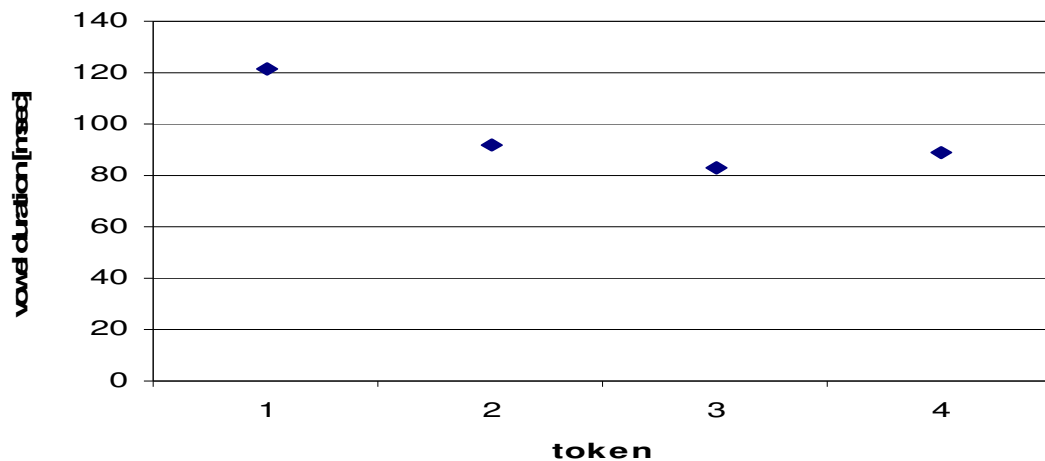


Figure 6. Vowel duration.

²⁹ The recordings in the two experiments were made with the help of a digital voice recorder Olympus VN-120 PC. The recording was made in the High Quality of voice mode, the recorder was equipped with an Electret Condenser Microphone (monaural).

Figure 6 illustrates the distribution of articulatory effort in terms of parameter 1, i.e. vowel duration. The prediction 1 is thus confirmed by the first parameter: the articulatory effort (duration) decreases along with the subsequent tokens of the word *web* used in the same context. The prediction 2, however, finds little support in the pilot study data: the articulatory effort increases in the final token of the word in question which appears in a different context only to a slight degree. The duration of the fourth token is shorter than expected.

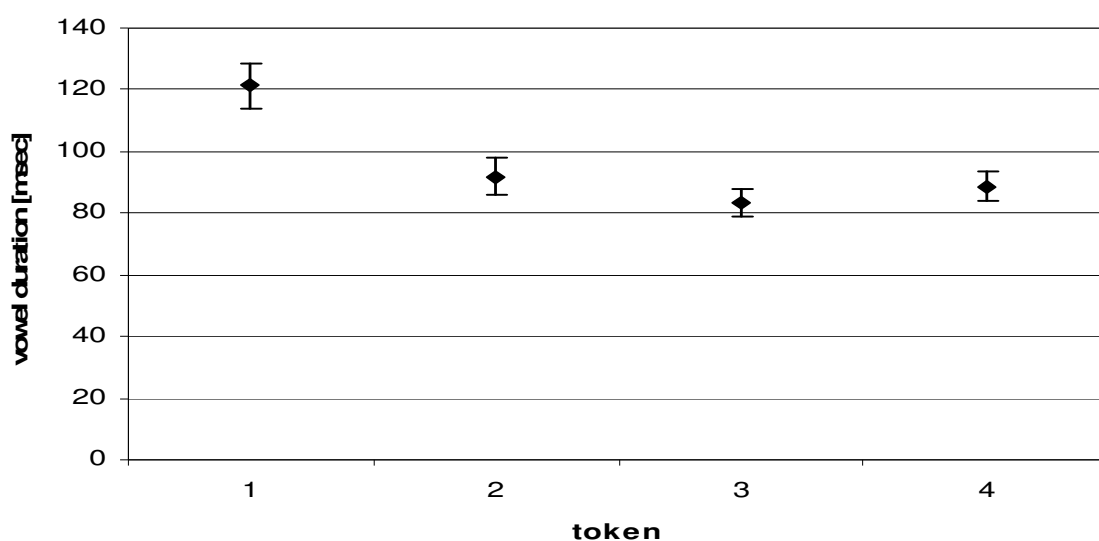


Figure 7. Confidence interval for vowel duration.

Figure 7 illustrates the confidence interval for vowel duration. The observed tendency is established with the 90 per cent probability as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable and do not occur due to a measuring error.

Next, the parameter 2 (vowel height) is discussed.

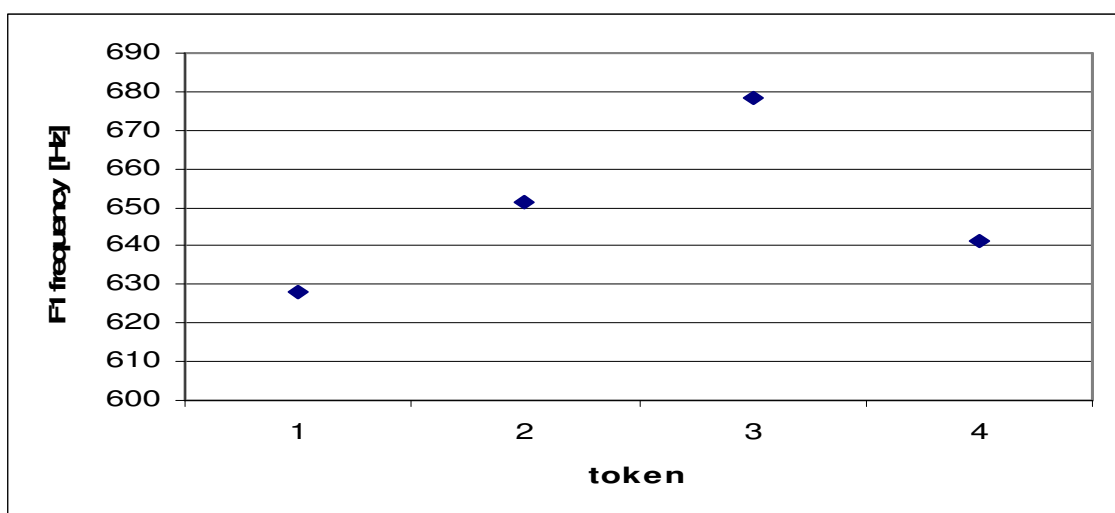


Figure 8. F1 frequencies.

Figure 8 does not inform about the vowel height in the subsequent tokens directly, therefore Table 4 was calculated as a reverse of F1 frequencies since there is an inverse relation between the height of a vowel and frequencies (the higher the formants, the lower the vowel).

Table 4. Vowel height.

mean	611,56	628,04	656,48	635,55
height	1	0,9738	0,9316	0,9623

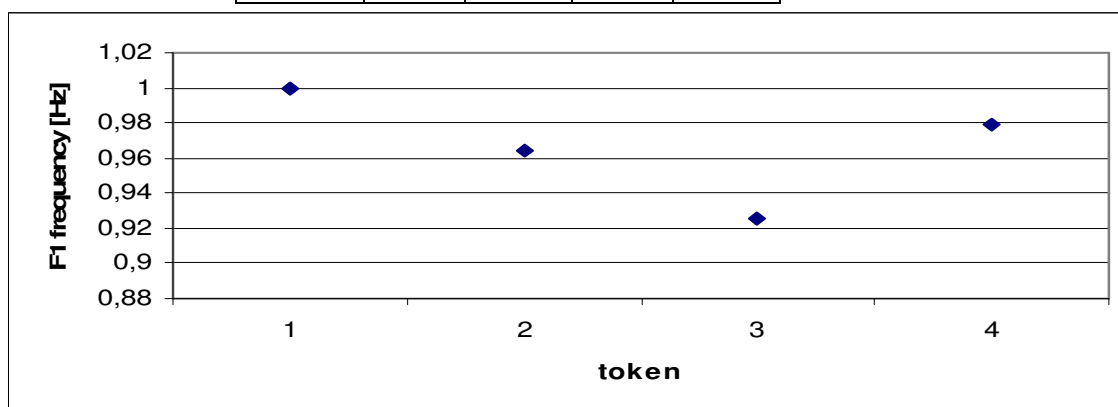


Figure 9. Vowel height.

Figure 9 illustrates the distribution of articulatory effort in terms of parameter 2, i.e. vowel height. The prediction 1 is thus confirmed by the second parameter: the articulatory effort (vowel height) decreases along with the subsequent tokens of the word *web* used in the same context. The prediction 2 finds more support in the pilot

study data as compared to parameter 1: the articulatory effort increases in the final token of the word in question which appears in a different context to a considerable degree, reaching almost the height of the first token.

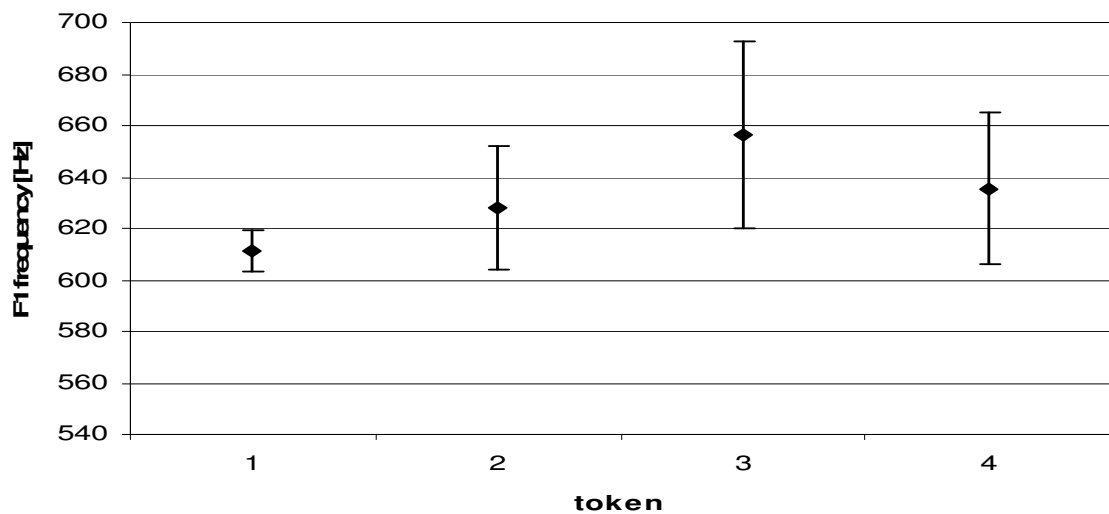


Figure 10. Confidence interval for Figure 8.

Figure 10 illustrates the confidence interval for the vowel height. The intervals do overlap to a significant degree, therefore the results obtained by means of parameter 2 (vowel height) are unreliable. The figure indicates that the differences in F1 values are not systematic, but rather might result from a measuring error. For instance, the confidence interval for the first token overlaps totally with the one for the second token. Thus, it can be assumed that there are no significant differences in vowel height. The figure also demonstrates that parameter 2 (vowel height) proved to be uninformative with respect to the overall effort.

Next, the parameter 3 (vowel position) is discussed.

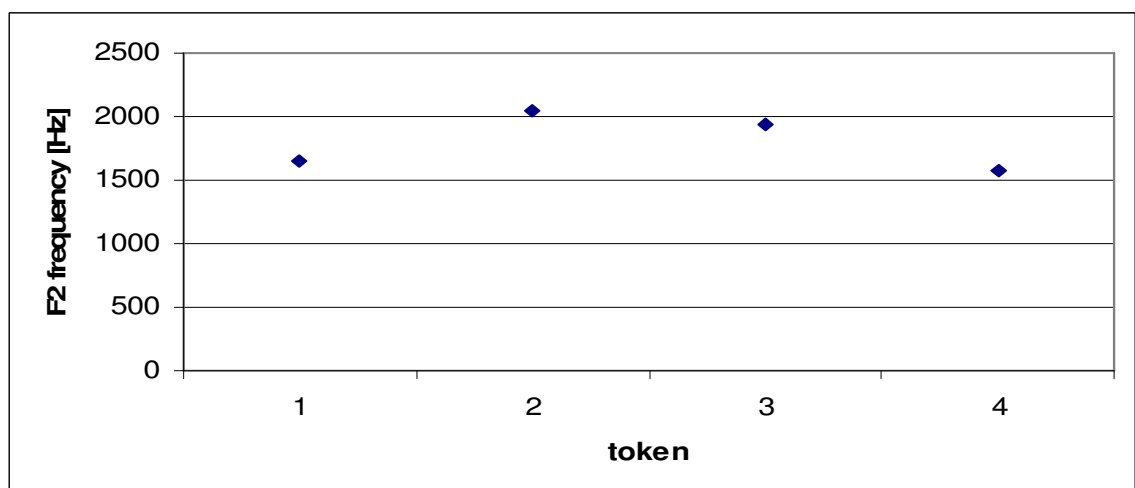


Figure 11. F2 frequencies.

Figure 11 illustrates the distribution of articulatory effort in terms of parameter 3, i.e. vowel position. There is an inverse relation between the backness of a vowel and frequencies (the lower the formant, the more back the vowel is). At first glance, neither prediction 1 nor 2 are confirmed by the third parameter. If one assumes, however, that /e/ is a front vowel, the subsequent tokens should be more and more retracted as from the front vowel /e/ to the central vowel /ə/ the tongue should have been retracted. Figure 11 shows an increase in frequencies, which means that the tokens are actually advanced. If one assumes that /e/ was velarized due to the influence of /w/, then advancement of subsequent tokens actually confirms the predictions.³⁰ There is a decrease in articulatory effort in the first three tokens, because the vowel centralization has occurred via its advancement from the velar place of articulation /w/, as stipulated by the prediction 1. There is an increase in articulatory effort in the final token as the tongue retracted towards the labial area (in order to articulate /b/, which indicates that prediction 2 finds support. The velarization assumption holds true only if F3 frequency shows an influence of labialization on /e/ (as /w/ is labio-velar).

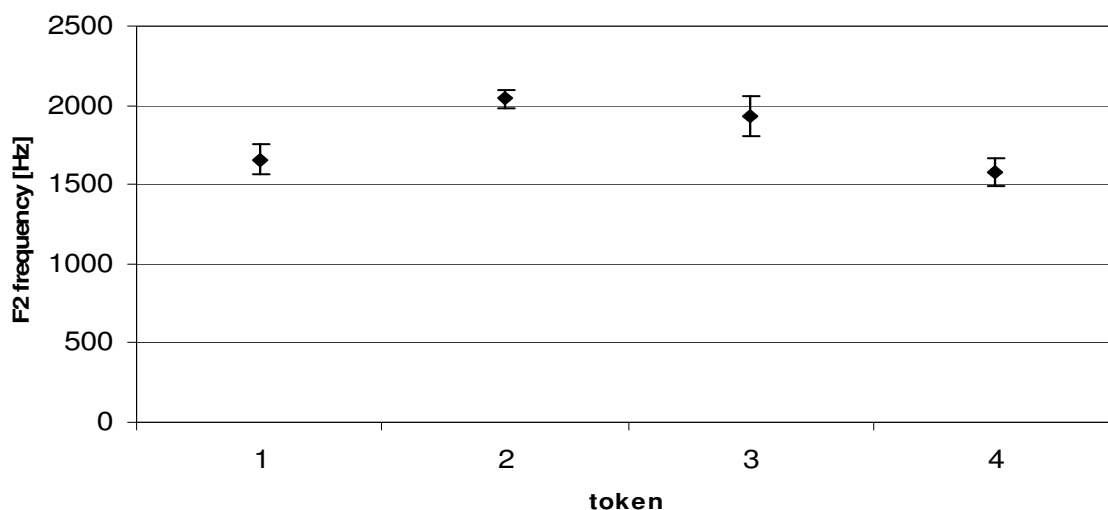


Figure 12. Confidence interval for Figure 11.

³⁰ In the word such as *web* the boundaries between the semivowel and the vowel are extremely fuzzy and hard to localize since semivowels are acoustically vowels. Consequently, it was not possible to compare the values of the formants for /e/ in isolation against the values of /e/ in the /w/ context. Therefore, the influence of /w/ was assumed due to the fact of the immediate vicinity and the concern that the measurements made for /e/ might have not been separated completely from these of /w/.

Figure 12 illustrates the confidence interval for Figure 11. The observed tendency is established with the 90 per cent probability as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable.

Additional measurement concerning the influence of /w/ on /e/ was made. In order to verify the velarization effect stipulated in the analysis of the parameter 3 (vowel position), the degree of lip rounding by means of F3 values is discussed. The velarization effect means that if the vowel /e/ was velarized (*web*), the centralization occurs via the advancement of the tongue, not its retraction.

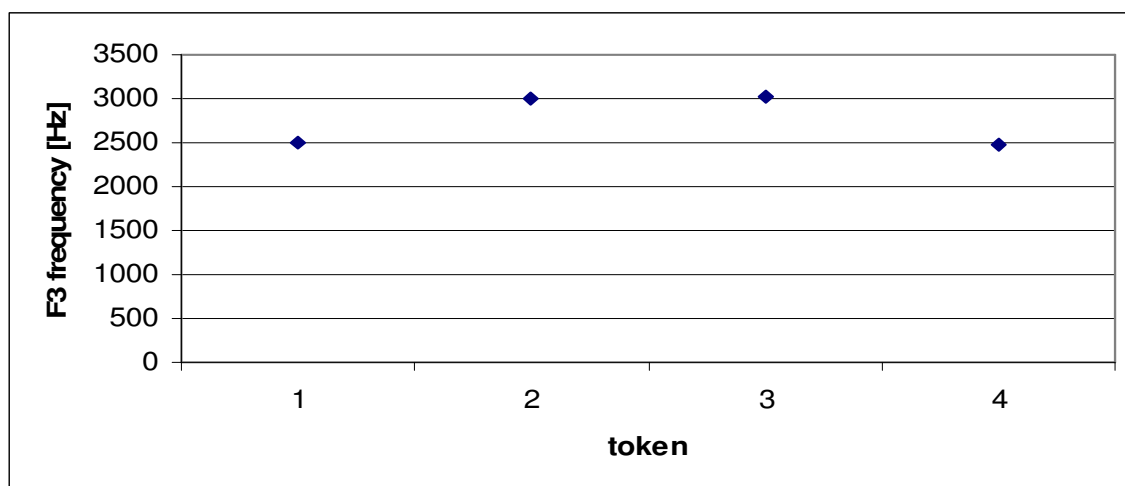


Figure 13. F3 frequencies.

Figure 13 illustrates the distribution of articulatory effort in terms of F3. F3 indicates the degree of lip rounding as the formants lower. In general, as sounds become more rounded, the frequencies of F2 and F3 decrease. Thus, the relatively low frequency of token 1 indicates an increase in articulatory effort (velarization due to /w/), a decrease in tokens 2 and 3 and again an increase in token 4 (an additional gesture in terms of labialization due to /b/). Besides, lip rounding is an extra gesture and increases the effort.³¹ The prediction 1 is confirmed by the parameter 3: the articulatory effort (lip rounding) decreases along with the subsequent tokens of the word *web* used in the same context. The prediction 2, however, also finds support in the pilot study data: the articulatory effort increases in the final token of the word in question which appears in a different context to a considerable degree, reaching the level of the first token. Figure 13 is a piece of evidence corroborating the velarization prediction and presents a similar

³¹ On one hand, coarticulation might be viewed as less effort as the gestures of /w/ and /e/ overlapped. On the other hand, /e/ is an unrounded vowel and the influence of /w/ adds to it an extra gesture which /e/ has not had before.

tendency as remaining the parameters. If F3 is lowered, the vowel is front. This advancement is due to the effect of velarization.

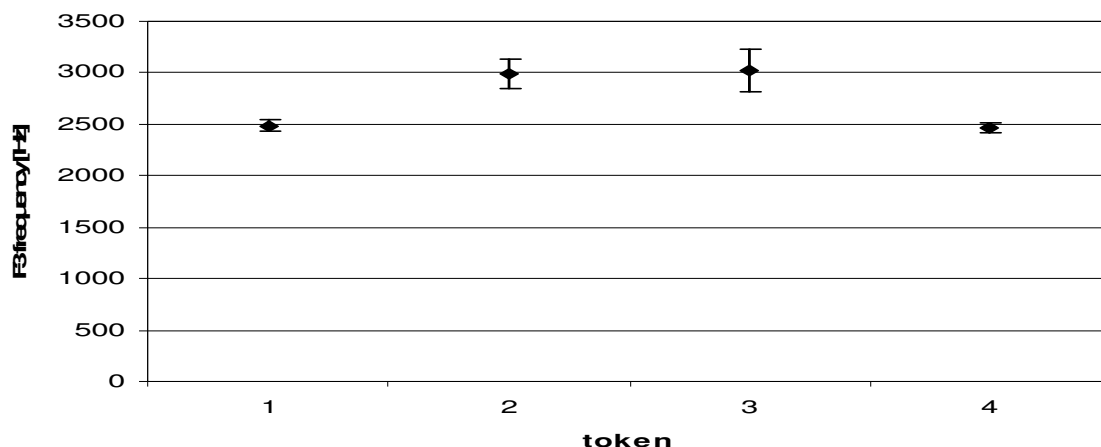


Figure 14. Confidence interval for Figure 13.

Figure 14 illustrates the confidence interval for Figure 13. The observed tendency is established with the 90 per cent probability as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable.

Next, the parameter 4 (obstruent devoicing) is discussed.³² This parameter consists in two factors, i.e. voicing of duration and closure of duration. Unfortunately, it was not possible to establish closure voicing due to the degree of background noise in the recordings which may have amplified the overall voicing throughout the recordings.³³ Therefore, devoicing was established exclusively on the basis of closure duration as closure voicing had to be abandoned due to technical obstacles. Duration of closure derives from basic articulatory settings. The first stage of obstruent production, the closure/approach stage, is a period of silence. The longer the period lasts, the more devoiced an obstruent is. In order to measure the degree of devoicing properly, the voiced obstruent /b/ was compared to its voiceless counterpart /p/ with respect to closure duration. For the sake of comparison, the word containing /p/ in the same

³² Devoicing is used here in the sense of the partial voicing category which corresponds to voicelessness if one adopts the three categories (fully voiced, devoiced, voiceless) instead of two (voiced, voiceless).

³³ In fact, Praat analysis of closure voicing may not be extremely reliable since in order to investigate voicing properly, one needs a laryngograph and an airstream velocity measurement device (Dafydd Gibbon, p.c.).

phonetic environment as /b/ was embedded in the text (*gap in earnings*). Then, the duration of closure for /p/ was included in the graph as the token number 5.

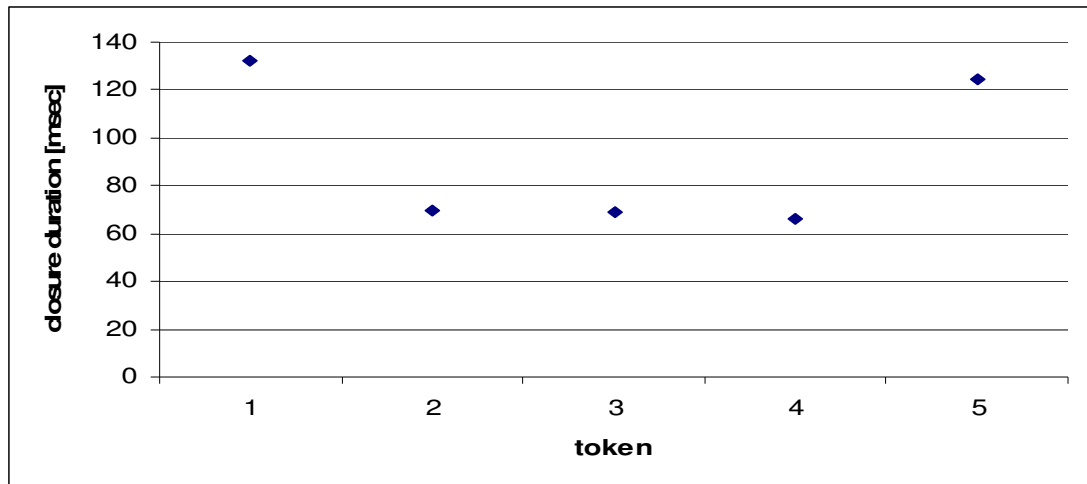


Figure 15. Closure duration.

Figure 15 illustrates the distribution of articulatory effort in terms of parameter 4, i.e. obstruent devoicing. Prediction 1 is thus confirmed by the first parameter: the articulatory effort (duration) decreases along with the subsequent tokens of the word *web* used in the same context. Prediction 2 finds the greatest support in the pilot study data among all parameters: the articulatory effort increases in the final token of the word in question which appears in a different context almost to the level of the first token. The duration of the closure lasts longer in the case of devoiced sounds and shorter for their voiced counterparts. Thus, the first token /b/ was clearly devoiced and its duration was almost the same as that of /p/ (the fifth token). Devoicing requires more effort, thus, tokens 1 and 4 were hyperarticulated. Tokens 2 and 3 were hypoarticulated and displayed the closure duration which was two times shorter than that of hyperarticulated tokens.

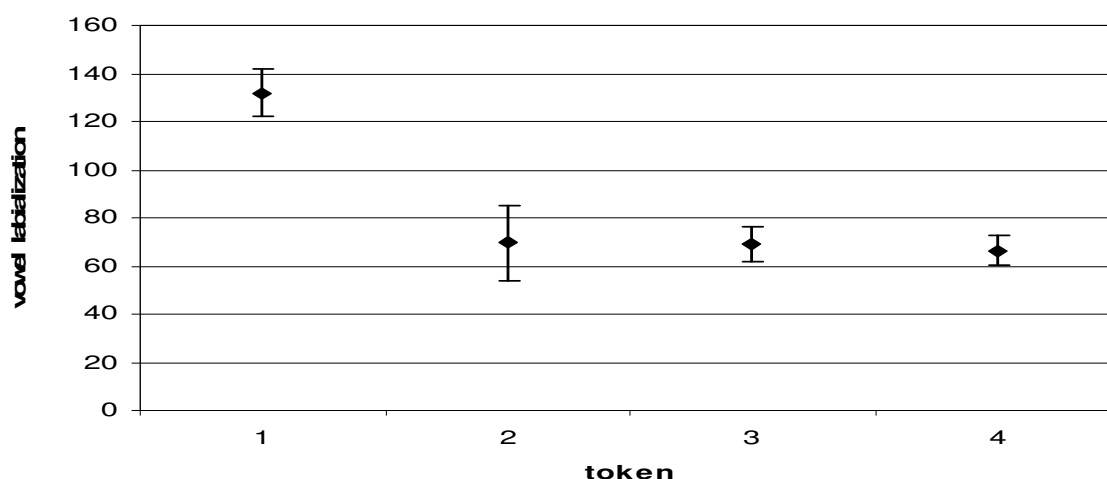


Figure 16. Confidence interval for Figure 15.

Figure 16 illustrates the confidence interval for Figure 15. The observed tendency is established with the 90 per cent probability as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable.

Next, the results obtained for parameter 5 (spirantization degree) are discussed. This parameter is the measure whether an obstruent in question was realized as a stop (more effort) or as a spirant (less effort). Thus, the higher the spirantization degree, the less effort is associated with the token. As far as the procedure of spirantization is concerned, the measure proceeded in few steps. Step one consisted in measuring both the durations of the preceding and the following vowel. The vowel following the obstruent /b/ was called V2, whereas the preceding vowel was called V1. In step two, the influence exerted by a stop on neighboring vowels was assessed. Stop has a short attack for vowel onset, whereas fricatives must develop and this process is time consuming. In Praat, the regularity and height of the vowel amplitude provides relevant information. The time in which the amplitude reaches the highest point (rise or fall, depending if the vowel preceded or followed the obstruent) was longer for spirants (about 92 seconds) and shorter for a stop (about 52 seconds). Step three consisted in calculating the ratio of the rise/fall duration for stop /b/ to the preceding and following vowel duration. This showed the consonant-vowel transition ratio and was given as percentage. The higher the percentage, the higher the spirantization degree as it means that transition took more time.

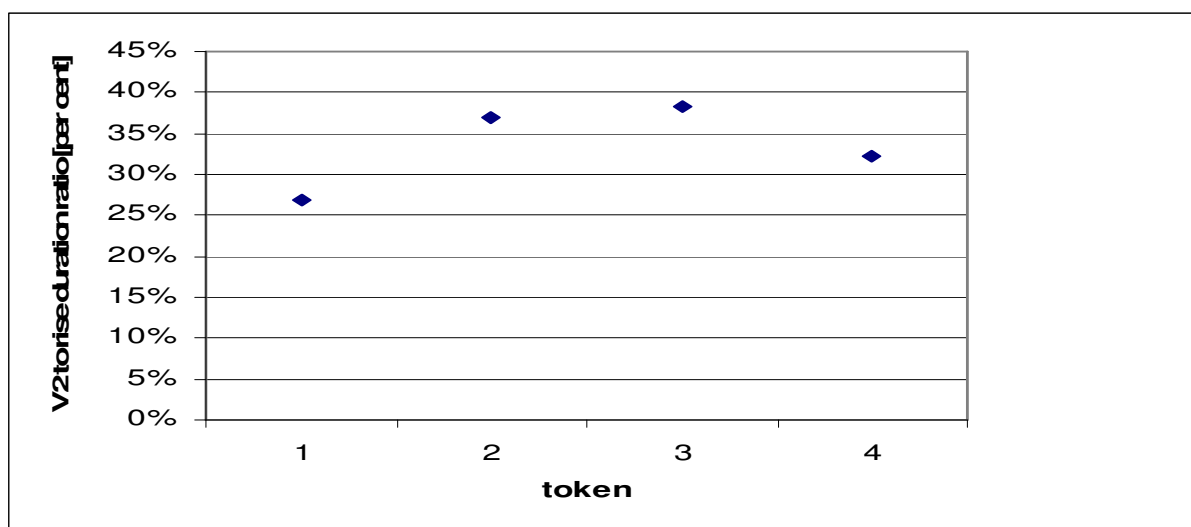


Figure 17. Ratio of rise to total duration for the following vowel.

Figure 17 illustrates the distribution of articulatory effort in terms of parameter 5, i.e. spirantization for the vowel following the stop. Prediction 1 is thus confirmed by the first parameter: the articulatory effort (duration) decreases along with the subsequent tokens of the word *web* used in the same context as the tokens are clearly spirantized. Token 1 had more features of a stop than tokens 2 and 3.

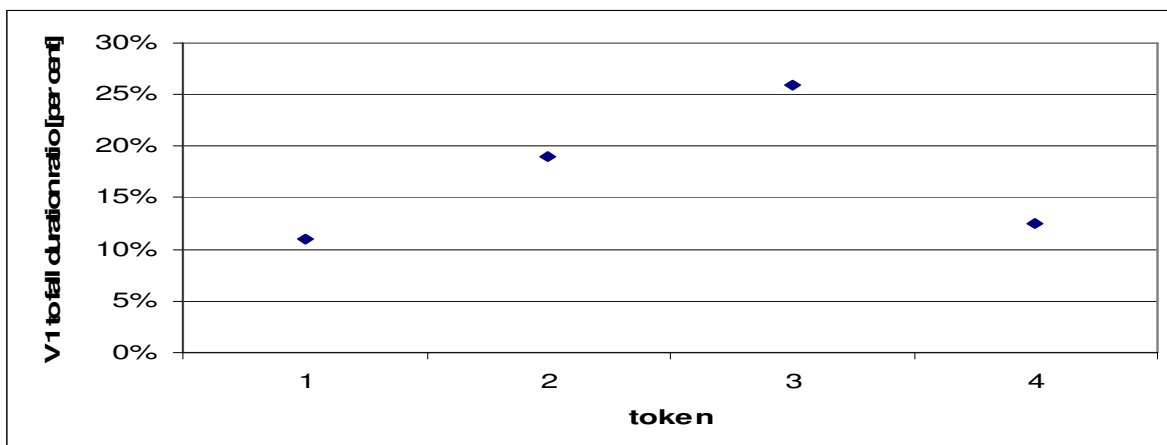


Figure 18. Ratio of fall to total duration for the preceding vowel.

Figure 18 illustrates the distribution of articulatory effort in terms of parameter 5, i.e. spirantization for the vowel preceding the stop. Prediction 1 is thus confirmed by the first parameter: the articulatory effort (duration) decreases along with the subsequent tokens of the word *web* used in the same context as the tokens are clearly spirantized. Prediction 2 finds the greatest support in the pilot study data among all parameters: the

articulatory effort increases in the final token of the word in question which appears in a different context as token 4 is less spirantized than the hypoarticulated tokens.

To summarize the results of the first experiment, three major conclusions can be drawn. First, the speaker behaves according to the least effort principle in managing his articulatory effort. The subsequent tokens used in the same context were more and more hypoarticulated because they were known to the speaker. Prediction 1 (tokens 2 and 3 used in the same context are hypoarticulated) is verified by parameter 1 (vowel duration, parameter 3 (vowel position), parameter 4 (final obstruent devoicing in terms of closure duration) and parameter 5 (the degree of spirantization) (parameter 5). Only parameter 2 (vowel height) failed to verify the prediction 1 due to the fact that the results obtained by its means were unreliable (Figure 10). Secondly, the semantic context only seemingly affects the changes of articulatory effort. Prediction 2 is partly verified since parameters produced mixed results. In fact, only vowel position and closure duration support it. Therefore, a question arises whether an increase in effort in the final token was indeed induced by the change of the semantic context. Consequently, a parallel experiment complementing the first experiment was carried out, with the fourth token of the word *web* used in the same context. The text of the parallel experiment is as follows:

The world wide web offers numerous resources. I use the world wide web in my work. I often used the world wide web in school assignments. For example, the world wide web is going to be useful because I need information about the Watergate scandal for my semester paper.

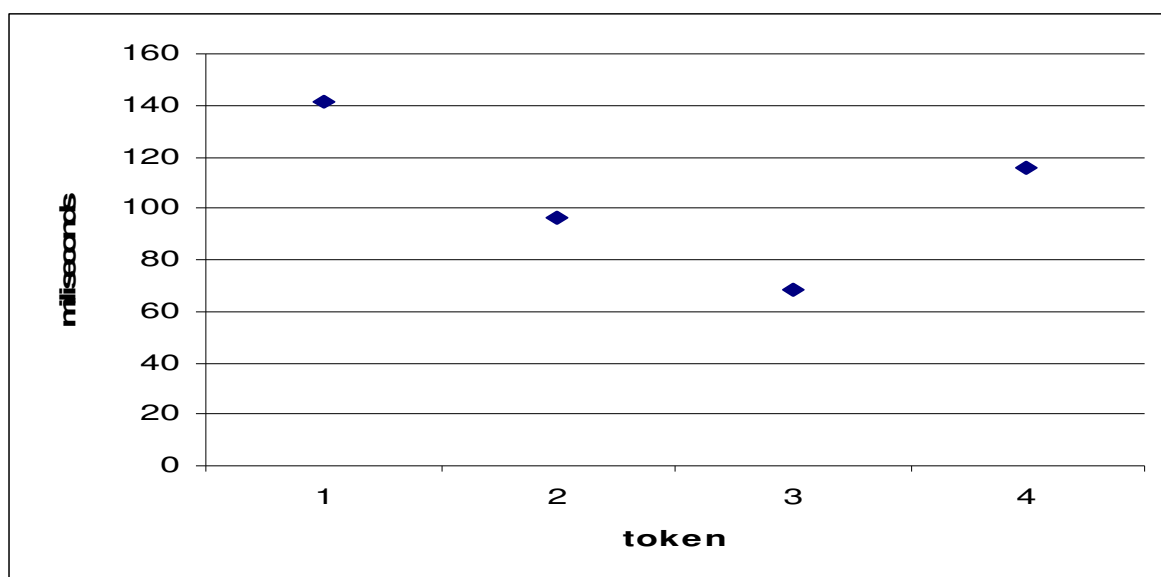


Figure 19. Vowel duration in tokens used in the same context [msec].

Figure 19 illustrates vowel duration in tokens used in the same context. There is a decrease in effort in tokens 2 and 3, while token 4 shows an increase in effort by an increase in vowel duration. The result obtained by the final token runs counter to the prediction that in the same semantic context, all the subsequent tokens should be hypoaarticulated. The figure is compared with Figure 6, which illustrates vowel duration with the change of the context:

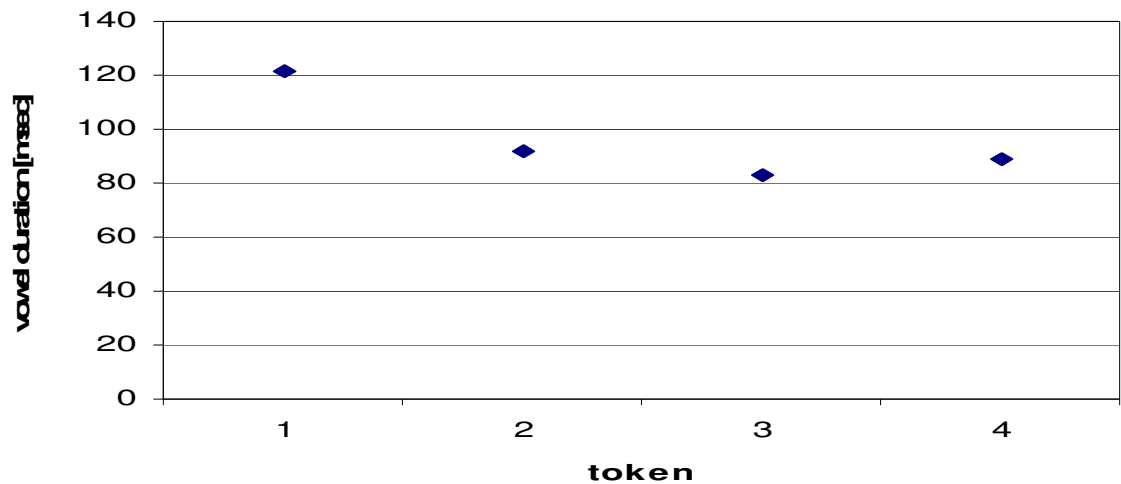


Figure 20. Vowel duration.

Therefore, the comparison demonstrates that regardless of the semantic context (the same or different), there is an increase of effort in the final token. The results indicated by the figure raise significant doubts about the validity of the prediction 2.

4.4.2. The results of the second experiment

The results obtained by means of the first experiment allowed to draw the conclusion that an increase or decrease in articulatory effort does not depend on a change in semantic context. Prediction 1, stipulating a decrease in effort in subsequent tokens remains in force. Moreover, it was observed that the first token invariably demonstrated the greatest effort. Thus, it was decided to abandon the semantic change (prediction 2 of the first experiment) and to explore prediction 2 in more details. Consequently, the predictions of the second experiment were reformulated as follows: The predictions for the second experiment are as follows:

Prediction 1: the first, introductory token should be hyperarticulated and involve the greatest effort of all tokens since new information is given and the speaker manages effort in order to be well understood.

Prediction 2: the subsequent tokens of the word should be hypoarticulated since the information is already known by the listener and no need to make effort arises.

Consequently, the second experiment was conducted. Its text is as follows:

The key player was absent in that match. I was sorry for his team because the key player has always scored most goals. Unfortunately, the key player's team lost the championship. It proved that the key player was indispensable.

As far as the effort parameters are concerned, it was observed that in the first experiment the five parameters equally contribute to the overall effort. Vowel duration, vowel position, vowel height, devoicing and spirantization of obstruents proved equally representative as they showed the same tendency (Figures 6, 11, 15, 17 and 18 respectively). Since all the remaining parameters provide an equal contribution to effort measure, suffices it to select only one of them. Moreover, the word *web* analyzed with respect to effort management and by means of the parameters turned out to be especially problematic for technical measurements as it was difficult to localize the boundaries between the end of the semivowel /w/ and the onset of the vowel /e/. Therefore, it was decided to place the word in question in the context of two voiceless stops /k/ and /p/ for a more sharp contrast, as in the phrase *key player*. Therefore, for the sake of the second experiment only one parameter, i.e. vowel duration was selected in order to verify the new, modified predictions.

For the sake of verification, twelve native speakers of English recorded the text of the second experiment. According to Figure 4 each speaker read the text seven times. Next, the means was calculated for every speaker. Then, the means of individual means was calculated and presented graphically:

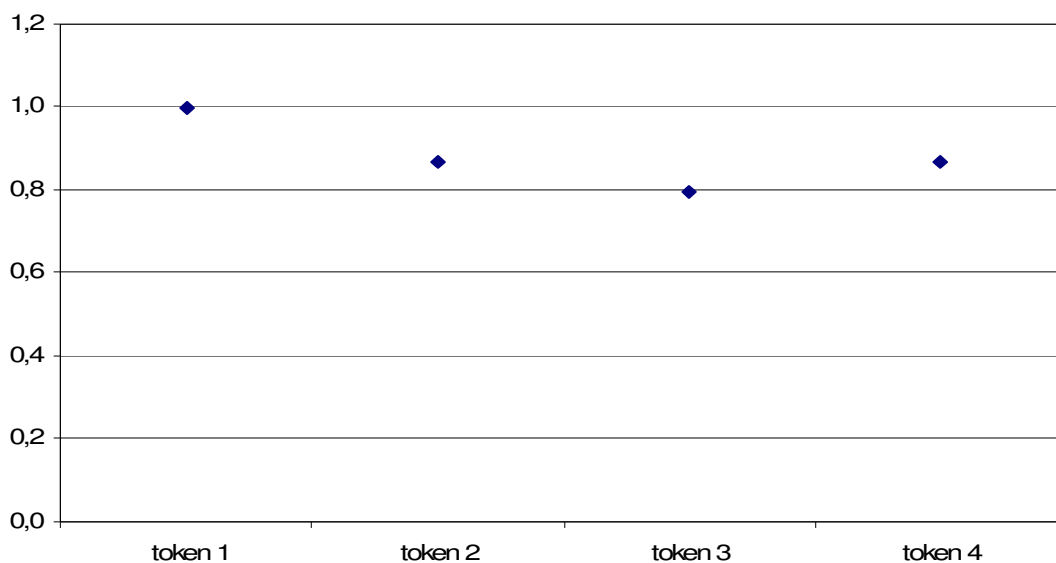


Figure 21. Vowel duration.

Figure 21 illustrates the duration of the vowel /i/. Prediction 1 is verified as effort is the highest in the introductory token. Prediction 2 is also verified, effort decreases with the second and the third tokens.

Table 5. Vowel duration in subsequent tokens of *key* [msec] (means of 12 speakers).

speaker	token 1	token 2	token 3	token 4
1	1	0,92	0,82	0,92
2	1	0,87	0,76	0,84
3	1	0,85	0,79	0,89
4	1	0,87	0,88	0,83
5	1	0,81	0,75	0,79
6	1	0,92	0,80	0,98
7	1	0,78	0,67	0,68
8	1	0,88	0,90	0,88
9	1	0,91	0,78	0,89
10	1	0,87	0,84	0,96
11	1	0,84	0,74	0,74
12	1	0,88	0,78	1,00
mean duration	1,00	0,87	0,79	0,87

Table 5 as well as Figure 21 represent the vowel duration for twelve speakers after normalization in terms of means. Therefore, having analyzed seven recordings made by twelve native speakers of English, the following conclusions can be made. The duration of token 2 was shorter by 13 per cent in relation to token 1, the duration of

token 3 was shorter by 21 per cent in relation to token 1 whereas the duration of token 4 was shorter by 13 per cent in relation to token 1.

As far as confidence intervals are concerned, it was observed that the speakers systematically shorten all the vowels in the subsequent tokens in the subsequent readings, probably due to fatigue, cf. table 6 where the values decrease with the tokens.

Table 6. Duration of the vowel /i/ in subsequent tokens of *key* [msec] (speaker 3).

Token 1	Token 2	Token 3	Token 4
70	57	51	57
65	49	46	60
61	52	41	48
57	50	46	50
60	45	49	43
53	50	49	55
58	56	52	61

In order to remedy this problem, normalization of the vowels was adopted, i.e. duration was compared to the first vowel in separate readings, and to establish confidence intervals for the normalized values. Besides, normalization will allow plotting the results obtained for all speakers on a single graph in order to compare the tendencies.

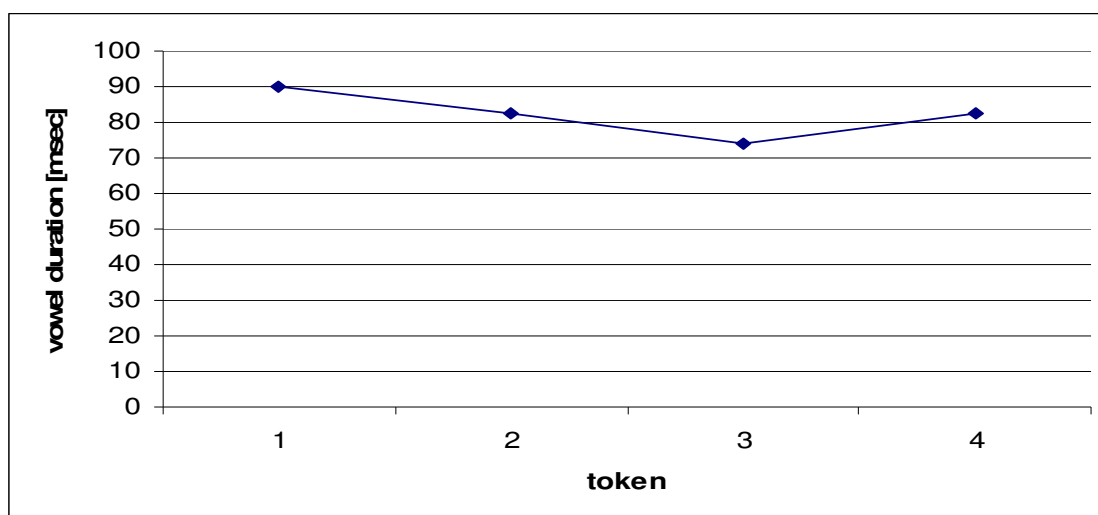


Figure 22. Duration of the /i/ vowel in subsequent tokens of *key* [msec].

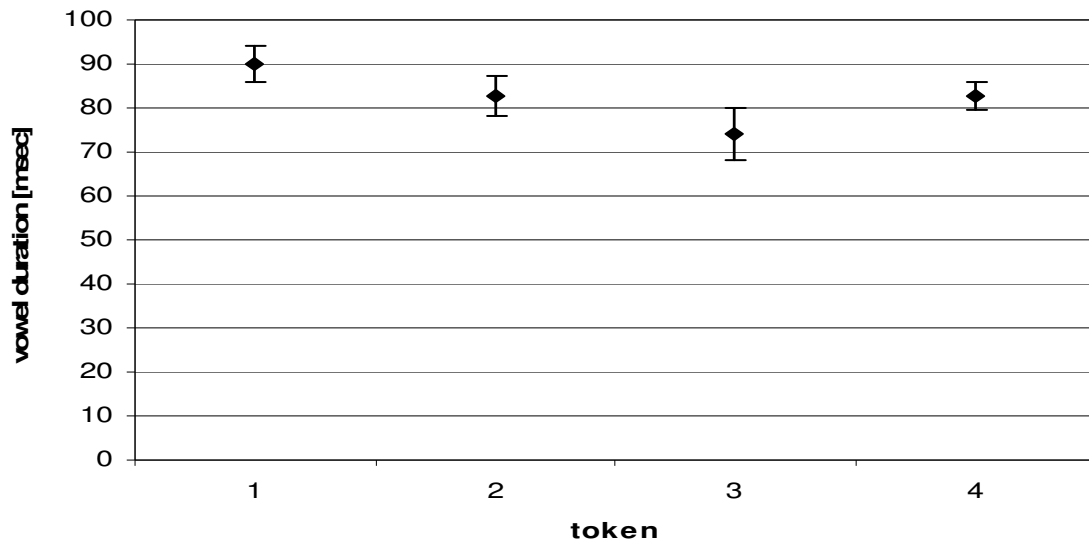


Figure 23. Confidence interval for Figure 22.

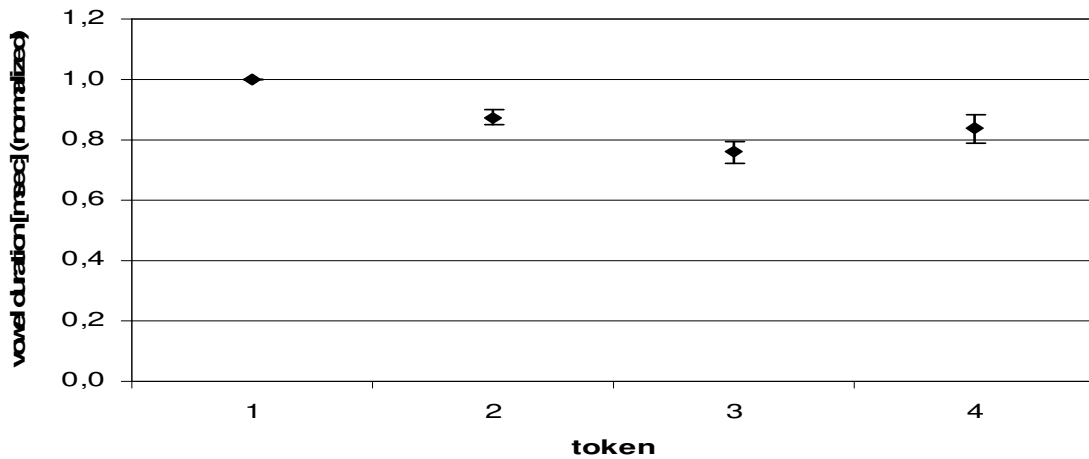


Figure 24: Normalized confidence interval for Figure 22.

Figures 24 illustrating the confidence intervals after normalization (i.e. the values of subsequent tokens in relation to the first one) shows a significant improvement in relation to Figure 23 and validates the obtained results statistically.

In terms of statistical significance of the second experiment (Figure 21), the analysis performed by means of confidence intervals (subsection 4.3.2.) produced the following results:

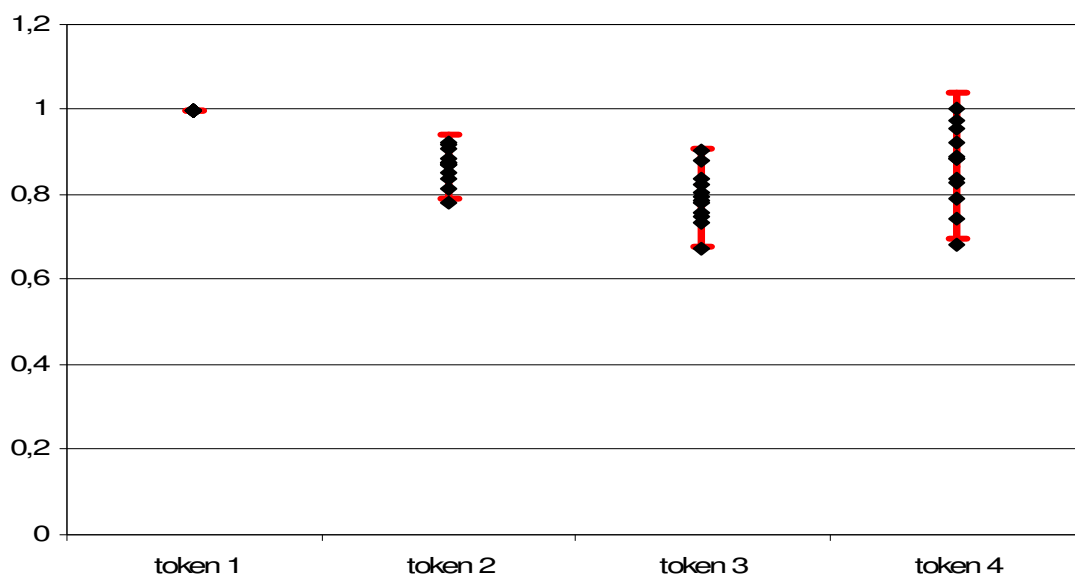


Figure 25. Confidence interval for Figure 21.

Figure 25 illustrates confidence interval for Figure 21. Every dot of Figure 25 represents seven recordings of speaker 1, speaker 2, speaker 3, speaker 4, speaker 5, speaker 6, speaker 7, speaker 8, speaker 9, speaker 10, speaker 11 and speaker 12. Figure 25, however, cannot validate obtained results statistically as it has been argued above that normalized values should be taken into account. Without normalization, Figure 25 shows only confidence intervals for variability in vowel duration, or for individual durations of speakers.

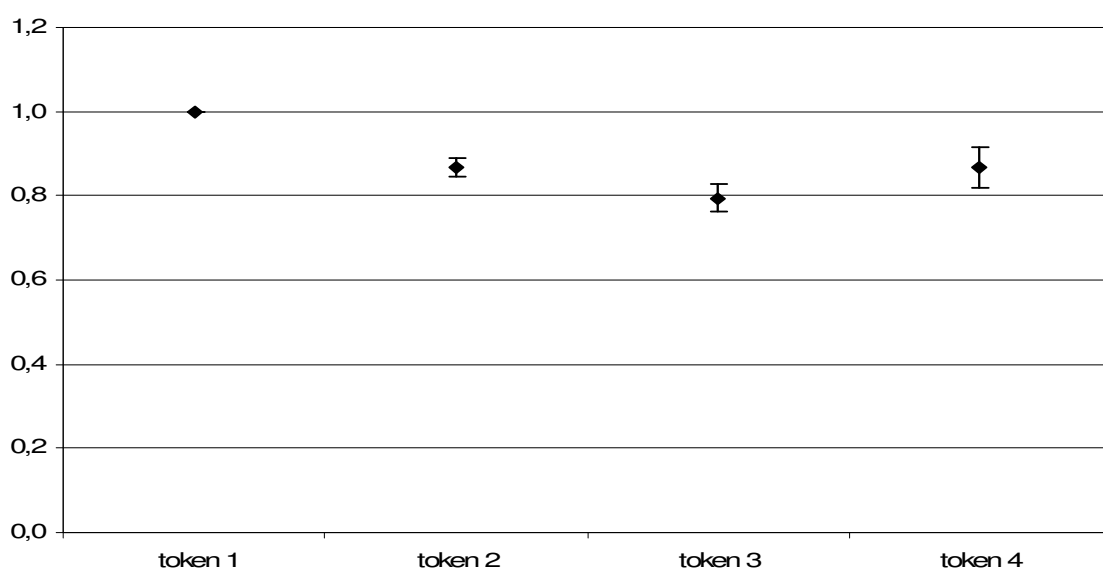


Figure 26. Normalized confidence interval for Figure 21.

Figure 26 illustrates normalized confidence interval for Figure 21. Every dot of the Figure 26 represents confidence intervals for means of seven recordings of twelve native speakers. Token 1 has no confidence interval as it has been normalized to 1 so that each subsequent vowel duration can be calculated as per cent relation to token 1. Figure 26 beyond doubts validates obtained results statistically with high probability as the measuring error was established at the level of 10 per cent and the observed tendency is established with the 90 per cent probability as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable.

4.4.3. The attention hypothesis

Interpretation of the results obtained for the final token presented a problematic question whether the increase in articulatory effort should be identified with effort management or a mere increase in attention. The fact that acoustic distinctiveness rebounds in the final token, might be a consequence of effort management (hypothesis 1: repetitions are more costly) or arise due to attention which comes back. It is hypothesized that the attention fluctuation mirrors the bathtub effect.³⁴ Psycholinguistics, in particular the lexical search and processing theories stipulate the importance of *frames* which are the beginning and the end of a word. If the frames exist, processing is possible even if the middle part is incomplete or misplaced, which is known as the *bathtub effect* (Brown and McNeill 1966):

People remember the beginnings and ends of words better than the middles, as if the word were a person lying in a bathtub, with their head out of the water one end and their feet out the other. And, just as in a bathtub the head is further out of the water and more prominent than the feet, so the beginnings of the words are on average better remembered than the ends [...] people tend to recall the beginnings and ends of words they cannot otherwise remember (Aitchinson 1987: 119).

Therefore, it can be speculated that the final token was hyperarticulated because of the bathtub effect. The area of public speaking provides evidence supporting the claim that the bathtub effect is also present in performance. Many speakers claim that they concentrate at the beginning and at the end since they witness the revival of attention when they anticipate the end of speaking. In order to verify the attention hypothesis, the

³⁴ The idea is that the particular position of a body in a bathtub makes head and feet the most prominent and head even more than feet. This applies to memory for words.

second experiment's text was modified so that the sentence with *key player* was not the final sentence of that text but another sentence follows it:

The key player was absent in that match. I was sorry for his team because the key player has always scored most goals. Unfortunately, the key player's team lost the championship. It proved that the key player was indispensable. Moreover, the coach got fired because his team didn't make it to the finals.

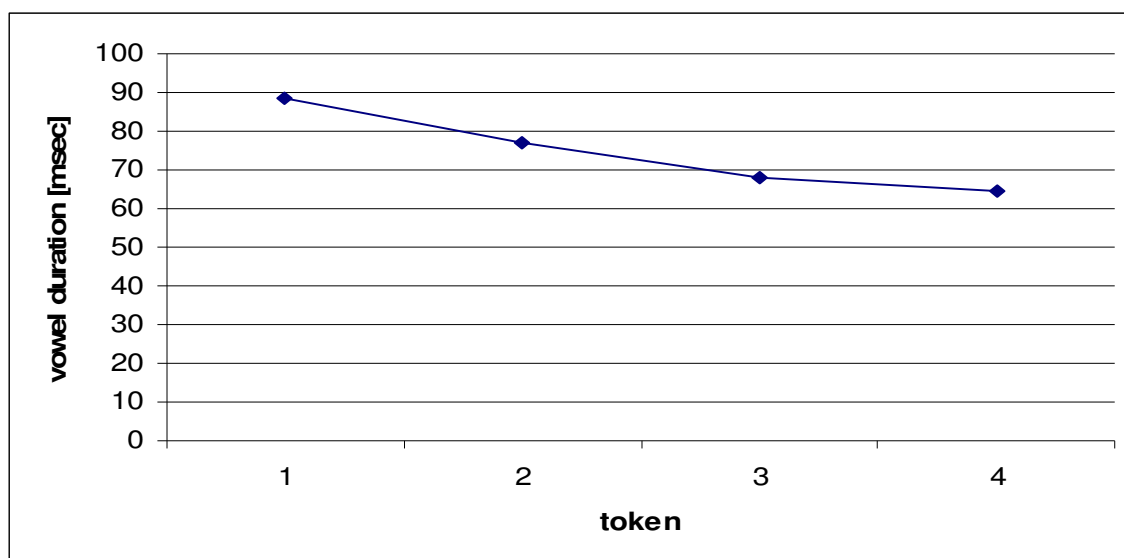


Figure 27. Duration of the /i/ vowel in subsequent tokens of key [msec] for the attention hypothesis.

Figure 27 illustrates the behavior of the final token in the text where it does not appear in the end of the text. The duration of the vowel in the final token is shorter relative to all the preceding tokens. Thus, the cyclical nature of attention, that it is high in the beginning and in the end like the bathtub effect, was confirmed.

4.4.4. The results of the third experiment

Since the results of the second experiment are obtained by means of a text which was developed on purpose in order to verify the predictions formulated in 4.4.2., it was considered necessary to verify the predictions independently of the second experiment and employ the method of triangulation (Jacob 1990). The method excludes a single research, source or experiment to check the validity of the findings, comparing the different ones. Therefore, the predictions confirmed by the second experiment were

additionally verified by the third experiment. It tested the predictions formulated in section 4.4.2. which are as follows:

Prediction 1: the first, introductory token should be hyperarticulated and involve the greatest effort of all tokens since new information is given and the speaker manages effort in order to be well understood.

Prediction 2: the subsequent tokens of the word should be hypoarticulated since the information is already known by the listener and no need to make effort arises.

The third experiment used the data obtained from a corpus of spoken English, MICASE (The Michigan Corpus of Academic Spoken English) (Simpson et al 2002). MICASE is a searchable corpus of academic speech events recorded at the University of Michigan, Ann Arbor. The sound files are accessible online (totaling 190 hours of recordings and 1,848,364 words), along with proofread transcripts. The files are categorized according to speech type events, such as class events (small and large lectures, discussion and lab sections, seminars, students presentations) and non-class events (including advising sessions, colloquia, dissertation defenses, interviews, meetings, office hours, service encounters, study groups, tours and tutorials). For the sake of the third experiment, the non-class speech events such as office hours in biology, poetry and economy were analyzed.³⁵ The idea behind the selection of office hours was to provide an informal, relaxed speech environment. According to MICASE website, office hours are “held by faculty or graduate student instructors in connection with a specific class or project” (<http://www.lsa.umich.edu/eli/micase/ATTRIB.html>). Moreover, the speech event type such as office hours in different subjects introduces more variety into researched words. Next, the relevant transcripts were searched for words which were repeated. The first mention of a word was tagged as token 1, whereas the repetition was tagged token 2. The list of thirty repetition words collected from MICASE is as follows:

³⁵ The duration of the analyzed sound files totaled to 4 hs 43 minutes.

Table 7. Repetition words from MICASE (office hours).

	Biology office hours	Poetry office hours	Economy office hours
1	frequency	sestina	plugged
2	allales	elegy	specific
3	sex-linked	poetry	point
4	thirteen	sentence	question
5	homogametic	clause	price
6	genes	true	lecture
7	female	respect	function
8	sixteen	subject	income
9	barred	different	math
10	birds	sound	
11		the rhyme	

Since thirty words from four separate online sound files were selected, the measure of articulatory effort in terms of vowel duration could not apply as there were different, often polysyllabic words containing various vowels. Thus, the duration of the whole word measured in milliseconds was adopted as a measure of articulatory effort. Then, their duration was analyzed in Praat (Boersma – Weenink 2005). Next, the duration of individual words was normalized, calculated as means for thirty words and presented graphically in percentage ratio:

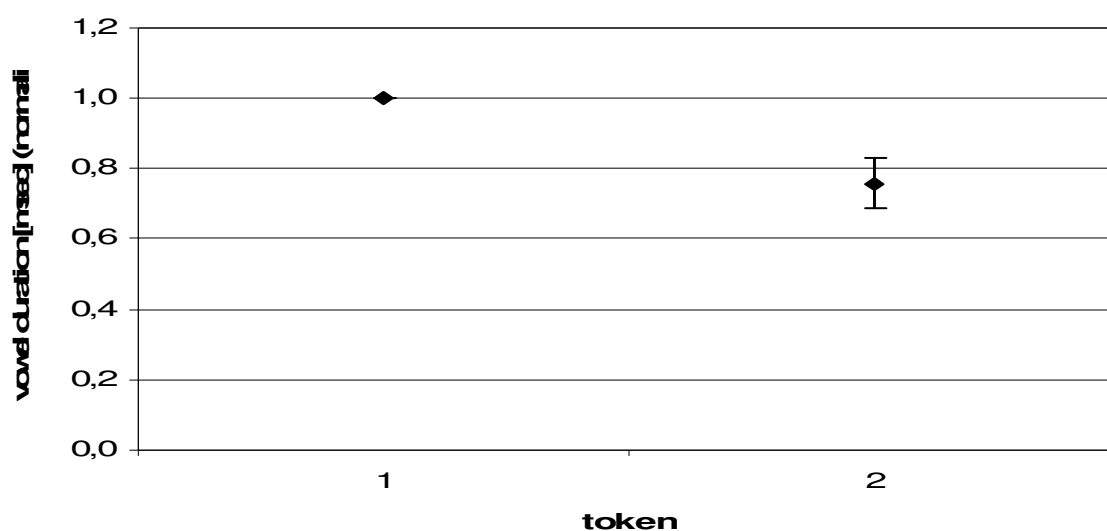


Figure 28. Duration of 30 words.

Figure 28 illustrates the duration of the thirty words after normalization in terms of means. Prediction 1 is verified as effort is the highest in the introductory token.

Prediction 2 is also verified, effort decreases with the second and the third token. Figure 28 also includes a normalized confidence interval for token 2 but not for token 1 as this one has been normalized to the number one in order to establish the percentage ration of token 2 to token 1. Moreover, the results of the third experiment establish the percentage ratio of token 2 to token 1 on the level of 24 per cent, which means that token 2 was on average 24 per cent shorter (hypoarticulated) than token 1. Figure 28 was calculated as means of thirty words in order to show an overall tendency since not all words from Table 7 behaved in the same way. In some cases, the difference between them was insignificant whereas in other the shortening was explicit.

In terms of statistical significance of the third experiment (Figure 28), the confidence interval is calculated at the significance level of 0,1. This means that the measuring error is 10 per cent. Since the significance level is 0,1 for the calculations, the degree of freedom is 0,90. The analysis performed by means of confidence intervals (subsection 4.3.2.) for the sample size of thirty recordings the t-Student coefficient being 1,699 (Table 3) produced the following results:

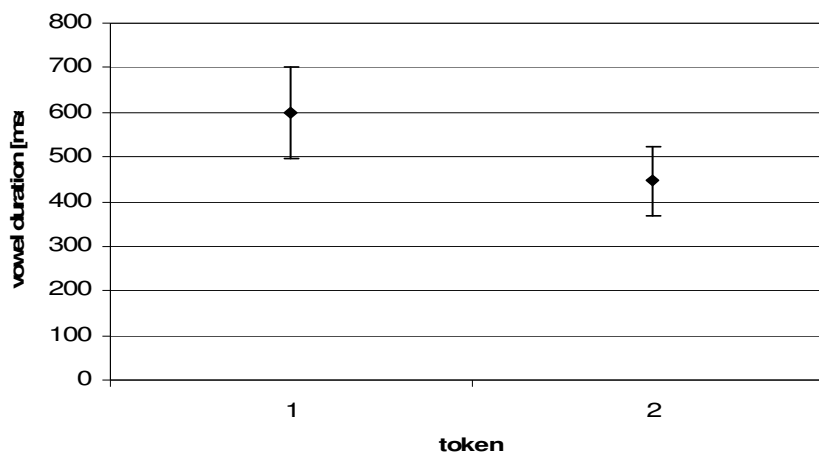


Figure 29. Confidence interval for Figure 28.

Figure 29 illustrates confidence interval for Figure 28. On the basis of Figure 29, it can be inferred that with 90 per cent of probability every token 2 of a word will be shorter from token 1. Figure 29 allows validating obtained results statistically with high probability as the measuring error was established at the level of 10 per cent as there is an insignificant overlap of the confidence intervals which suggests that the results are reliable.

4.5. Discussion of the results

The results of the second and third experiments verify the two hypotheses:

Prediction 1: the first, introductory token should be hyperarticulated and involve the greatest effort of all tokens since new information is given and the speaker manages effort in order to be well understood.

Prediction 2: the subsequent tokens of the word should be hypoarticulated since the information is already known by the listener and no need to make effort arises.

Figure 21 demonstrated that the principle of least effort, reformulated as effort management, is in operation. Moreover, the attention prediction claiming that attention and resulting effort is of cyclical nature, which means that attention rebounds in the final has been proposed and verified by Figure 27. Therefore, one can conclude that effort management plays a role in the relation between the speaker and the listener. Articulatory effort, measurable by parameters, is the negotiating platform between the two participants of the communication act and it is the speaker who takes into account his need to be understood. The speaker manages his/her effort accordingly, by hyperarticulating when necessary (e.g. in the introductory tokens) and by hypoarticulating when s/he judges that no need to be hypercorrect arises (e.g. subsequent tokens). Finally, effort management shifts the burden of articulation from the language internal properties such as semantics (the results of the first experiment excluded semantics as a trigger of effort changes) onto communicative situation (introductory token hyperarticulated, while the subsequent ones hypoarticulated). Not only do the two predictions constitute the effort management verified in a read text (the second experiment) but also by a corpus of spoken English (the third experiment).

The aim of the thesis was to establish a preference hierarchy with reference to articulatory effort. It was conducted on the basis of the three experiments which produced statistically significant results. The results allowed observing a tendency of the speakers to systematically shorten duration of the subsequent tokens (in particular, vowel duration in the second experiment and whole word in the third experiment) which can be called speakers preferences. The preferences can be graphically represented on the following scale:

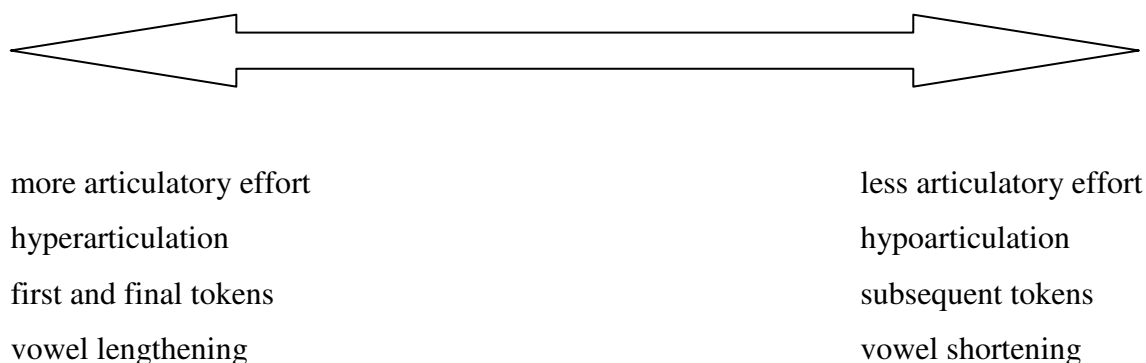


Figure 30. The scale of speakers' preferences with reference to articulatory effort.

Figure 30 represents the scale of speakers' preferences as far as their articulation is concerned. On one end on the scale there is more articulatory effort which demonstrates that the speakers preferred to hyperarticulate the first token which introduced the word in question as well as the final token in which articulatory effort rebounded (this is explained by the cyclical nature of attention hypothesis). On the other end of the scale, there is less articulatory effort which demonstrates that the speakers preferred to hypoarticulate the subsequent tokens. This scale also illustrates the operation of effort management which predicted and verified the preference of the speakers to expend more effort when it was considered necessary (the first and final tokens) and less effort when the word was already known and there no need arises to hyperarticulate.

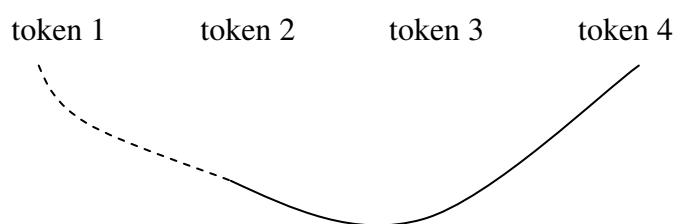


Figure 31. The hierarchy of speakers' preferences with reference to articulatory effort.

Figure 31 illustrates the speakers' preference on the basis of the second and third experiments. The dotted line indicates the preference in the third experiment (as there were two tokens) whereas the whole line demonstrates the preference for the second experiment (there were four tokens).

Since effort management has been therefore proven valid and operative in spoken language, it is considered necessary to check its operation in written language as

well. This will be demonstrated on the example of text messages adopted as an illustration of written language.

4.6. Effort management in text messages

Text messages, or SMSes (Short Message Service), constitute a special case of communication since it is restricted by the protocol which allows only 160 characters per single message, including spaces between words and by the presence of alphanumeric phone keypad. Furthermore, the vast majority of text messaging is carried out in order to make arrangements (Kuźminski 2004) and it proceeds along the dialog line as text messages typically involve a rapid exchange. Therefore, the above conditions which determine minimal form inevitably foster the use of economy strategies. This term denotes the variety of productive devices motivated by the urge to economize on space and words. In the literature, they are also referred to as contractions, abbreviations and consonant writing (Papen – Tusting 2006, Thurlow 2003, Segerstad 2002) or shortening devices such as clippings, phonetic respellings, capitalizations, letter homophones and number homophones (Lopez Rua 2006, Kul in press). Therefore, economy strategies employed in text messages can be legitimately viewed and analyzed as a manifestation of effort management due to the fact that limited space must be managed in order to achieve minimal form of maximal content. Moreover, the person who sends a text message must minimize the form so that the message is decodable for the receiver. Writing full forms of words, especially the polysyllabic ones, is effort consuming, thus the sender employs economy devices with a view of reducing the workload.

In order to exemplify the treatment of text messages as effort management, three studies were conducted. The first study investigated metaphonological devices in English text messages (Kul 2007). The second study analyzed ten examples of English text messages coming from various sources such as websites or press articles in an attempt at determining if the deletion of letters was regular, the general prediction being that text messages are decoded via the mediation of their phonemic representations (or via mental reading). The third study presented and discussed economy strategies in Polish text messages (Gibbon – Kul 2006). In the studies of English and Polish text

messages, the following economy strategies have been identified and they will be discussed as manifestation of effort management:

4.6.1. Omission of punctuation

Omission of punctuation allows economizing on characters in a highly circumscribed environment. Examples: a) *Ojciec powiada ze nie ma zadnego na zbyciu bezpiecznika. Jesli to sa wykrecane mozna zastapic tanio go zwyklym bez switcha ale najlepiej nowe nabyc* b) *Dobranoc kochanie ja czytam ksiazke od ciebie i popijam piwko zaraz pora spac bo tez jestem zmeczony a jutro chcialbym duzo zrobic caluski kochanie papapapa* c) *Jakby drzwi od klatki byly zamkniete to pusc strzalke bo nie mamy domofonu* d) *i nd sum1 2 lv me n care 4 if ur dat 1 pls lt me knw 2nite cuz i cnt w8 any mr.*

4.6.2. The use of small letter instead of capitals in proper names

The use of small letter instead of capitals in proper names affords writing convenience. The change from capital to small letters requires an additional operation which disturbs and prolongs writing. No effort is required when small letters are used at all times, thus the sender consciously eases his/her task of writing. Examples: a) *Po 17 na rondzie przybyszewskiego* b) *Ja koncze 18.45 na zamkowej to czekaj na mnie absoluta mozemy potem isc na ten koncert na matejki w jakims pubie* c) *Gdzie dostales ladefogeda?* d) *prostopadla do garbar* e) *Czekolada jest blizej solnej* f) *vry hapy nu yr.*

4.6.3. The lack of spaces between words and the use of capitals to mark word beginnings

The lack of spaces between words and the use of capitals to mark word beginnings is motivated by economy since spaces between words are also charged by the mobile phone operators. Thus, the extra space is used for writing and the word boundaries are marked with capital letters so that it is clear where a word begins. In terms of effort

management, it takes one stroke of the key to switch from small letter mode into the capital letter one instead of using the space every time a new word starts.

Examples:

a) *Hey! CouMnie?zDominikemGodzimySie(tymRazemOnZerwalBoNiePosluchalamGoIspalamUkolagiAleInnejMozliwosciNieMialamBoZalezaloMiByGdzisPojscINieZdazylamNaAutobusWzgDoKlenicy)jeszczeNiePracujeNaStaleIjestemWMCHOD.ZaadamemTesknieJakbymGoSzalenieMocnoKochalaAleNadalJestemKonsekwentnaWmilczeniu @.@pa*
b)

OstatniaSeriaPytanAkapityMajaBycNa2cm?JakSiePiszeRosyjskiSkrotOdNumerICzyToJestNapewnoTakZeWBibilografiiOsobnoPolakRuskiAWPracyWgDat? c)
IfYaMthWozNEBiGrUWdntHavNEFAcLft2Wsh d) *TonlyThngThtLOksGOdOnMeIsU.*

4.6.4. Emoticons

Emoticons are universal and create the possibility to express feelings and attitudes in an economical way, by means of a single symbol. The use of a single symbol replaces verbal expression of emotional content as it is shorter and less effort-consuming to write ;D than *just kidding*. Moreover, in many mobile phone models no need arises to produce the emoticons by hand as they constitute a separate set of symbols, they are simply inserted by the sender which saves time. Examples: :p, :D, :(, :), ☺.

4.6.5. Onomatopoeic expressions

Effort is managed via employing universally recognized onomatopoeic expressions standing for attitudes, feelings and emotional comments. Examples: *ehhh, buhaha, blabla, auuu, hehe, hihi, aaaa, ups, grr, booo.*

4.6.6. Replacement of the Polish characters with the ASCII ones

ASCII (American Standard Code for Information Interchange) stands for the character encoding which represents text in digital devices, e.g. computers, mobile phones etc. It

reflects digital bit patterns as glyphs/symbols of written language, enabling sharing, storing and processing all character-oriented information. ASCII is based on the English alphabet. It was observed that no text message from the Polish corpus used the Polish characters. The popularity of this particular strategy may be linked to a number of reasons. Certain models of mobile phones fail to display Polish characters. Consequently, the characters are replaced by the £ symbol or with the Greek ones. ASCII incompatibility results in the “window effect”, comparable to the one witnessed in computers text, e.g. *si*□. It is possible to retrieve the character from the context, but inserting it is a waste of time and mismanagement of effort. Another reason for replacement of Polish characters is lack of practice and training. Since a Polish character is associated with a given key as the subsequent (be it the fourth or the fifth) character, it requires precision to spot the right symbol. Obtaining Polish characters by means of skipping the other ones presents no problems for teenagers, but is frequently reported to be difficult by adults. Finally, there are no lexical competitors of the word such as *sie* ‘się’, *bedzie* ‘będzie’ or *sa* ‘są’, therefore, the possibility of confusion does not exist. In general, ASCII is used since it is more economical and as such, sufficiently illustrates effort management. Examples: *coz* ‘cóż’ (well), *palic* ‘palić’ (to some), *w ciagu* ‘w ciągu’(during), *sie* ‘się’(itself), *spoznie* ‘spóźnić’(I’ll be late), *zaplaca* ‘zapłaca’(they will pay).

4.6.7. Borrowings

They occurred in Polish text messages as the subjects of the third study have a decent or good command of English, therefore they use English words and expressions when they judge that the English word is shorter than the Polish one. No case was observed that the borrowed word was longer than the Polish equivalent. Borrowings of English words afford the possibility to economically manage space and time.³⁶ Examples: *h* ‘godzina’ (hour), *asap* ‘najszybciej jak można’ (as soon as possible), *gut news* ‘dobre wiadomości’, *happy new year* ‘szczęśliwego Nowego Roku’, *see you* ‘do zobaczenia’, *nxt* ‘następna’, *4 all* ‘dla wszystkich’, *sat* ‘sobota’, *sun* ‘niedziela’, *new* ‘nowy’.

³⁶ A case of borrowing from German was also encountered: *bany* ‘pociągi’ (trains),

4.6.8. Number homophones

Number homophones employ metaphonological phenomenon of graphemic-phonemic manipulation and their potential lies in the homonymy effect. It is used in order to manage effort so that parts of words are replaced with a number. Examples: *3maj sie*, *3m sie* ‘trzymaj się’ (take care), *2nite* (tonight), *2morrow* (tomorrow), *3dom* (freedom), *4ever* (forever), *4tun* (fortune), *m8* (mate), *gr8* (great), *w8* (wait).

4.6.9. Letter homophones

As above, homonymy serves as a vehicle to save time and effort, thus its effects enhance economy. Examples: *S*, *ska* ‘eska’ (text message), *b* for *be*, *m* for *am*, *n* for *an/and*, *o* for *oh* (*oh I see*), *UR* for *You are* or *your*, *c u* for *see you*, *y* for *why*.

4.6.10. Letter reduction

Letter reduction constitutes the most particular case of effort management. It can be explained by means of the semiotic principle of “figure and ground” (Dressler 1996). The principle “predicts that figures tend to be foregrounded, grounds to be further backgrounded” (Dressler 1996: 42). Thus, consonants can be compared to figures and vowels to grounds. In speech, consonants are likely to be preserved, whereas vowels are likely to be deleted. The third study of English text messages investigated letter deletion in English text messages. It concluded that the semiotic figure and grounds principle (Dressler 1996) was in force since consonants in initial and final positions are likely to be preserved and vowels are likely to be deleted. This also implies that letter reduction observes the idea of effort management to the biggest extent. The fact that consonants were not deleted, at least not in word initial and final positions, demonstrates that there are constraints on reductions which the sender of text messages must take into account. If most consonants were deleted, effort would be mismanaged and the receiver would not understand the message.

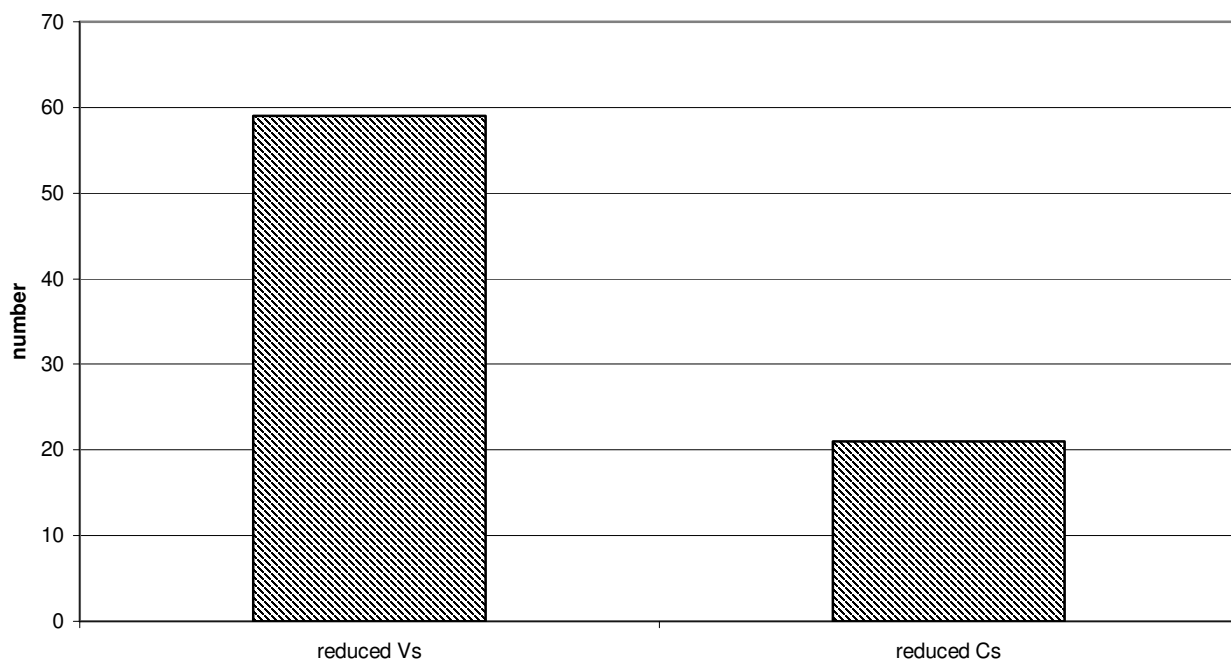


Figure 32. Proportion of deleted vowels to consonants

The values in Figure 32 are given in numbers where the two types of deletion are compared against each other. Figure 32 shows that as many as 59 vowels were deleted, whereas only 21 consonants were subject to deletion. The data shown in Figure 32 also illustrate the fact that in general vowels which do not carry the functional load are more likely to be deleted: more than twice as many vowels have been deleted in comparison with consonants. Examples: *pzdr* ‘pozdrawienia’, *kwrtlnch* ‘kwartalnych’, *tlkin* (talking), *vry* (very), *hrt* (hurt), *frm* (from).

4.6.11. Phonetic respellings

The effort required by lack of one-to-one correspondence between graphs and phonemes in English orthography is managed via phonetic respelling. In Polish, orthography is more iconic so that the only examples are special cases of respellings in which the English lexical items are phonetically represented with the Polish characters. Examples: *lawju* ‘kocham cię’ (I love you), *fak* ‘kurwa’ (fuck), *gut najt* ‘dobranoc’

(good night), *spicz* ‘wystąpienie’ (speech), *slit drims* ‘słodkich snów’ (sweet dreams). *nu* (new), *yaself* (yourself), *nite* (night), *coz* (because), *afta* (after).

4.6.12. Clippings

This economy strategy also effectively manages effort on the part of the sender, especially in Polish text messages as Polish has a high number of polysyllabic words (due to inflectional and derivational endings). Consequently, there is a tendency to reduce them to the first syllable. Examples: *coz* (because), *rach tel* ‘rachunek telefoniczny’ (phone bill), *min* ‘minuta’ (minute), *inf* ‘informacja’ (information), *dot* ‘dotyczący’ (regarding), *pozdro*, *pozdr* ‘pozdrawienia’ (greetings), *cze* ‘cześć’ (hello), *mam nadz* ‘mam nadzieję’ (I hope), *b* ‘bardzo’ (very), *spr* ‘sprawdzić’ (to check), ‘sprawdzian’ (test), *kom* ‘komórka’ (mobile phone), *syg* ‘sygnał’ (phone signal), *godz* ‘godzina’ (hour), *cz* ‘czy’ (if), *ew* ‘ewentualnie’ (or), *rozm* ‘rozmowa’ (talk), *dop* ‘dopiero’ (at least), *max* ‘maksymalnie’ (maximum), *narka* ‘na razie’ (bye), *sorki* ‘przepraszam’ (I’m sorry, excuse me), *powt* ‘powtarzać’ (repeat), *st* ‘stopień’ (degree), *trza* ‘trzeba’ (one needs to), *psych* ‘psychicznie’ (mentally), *fiz* ‘fizycznie’ (physically), *mo* ‘mocno’ (strongly), *do zob* ‘do zobaczenia’ (see you), *zal* ‘zaliczenie’ (credit), *dzieks* ‘dziękuję’ (thank you), *mikra* ‘mikroekonomia’ (microeconomics), *impra* ‘impreza’ (party), *odp* ‘odpowiedz’ (reply), *komp* ‘komputer’ (computer), *nr* ‘numer’ (number), *str* ‘strona’ (page), *pasi* ‘pasuje’ (suits), *bibl* ‘bibliografia’ (references), *art* ‘artykuł’ (article), *net* ‘internet’ (the internet), *zadz* ‘zadzwoń’ (to call), *po powr* ‘po powrocie’ (on coming back), *wiad* ‘wiadomość’ (message), *spozn* ‘spóźnić się’ (to get late), *dyska* ‘dyskoteka’ (disco), *pusc mi* ‘puść mi strzałkę’ (send me a signal), *pozost* ‘pozostawiliśmy’, *najlep* ‘najlepiej’, *zaj* ‘zajęcia’ (classes), *wcz* ‘wczoraj’ (yesterday), *czyt* ‘czytaliśmy’ (we read), *rob* ‘robiliśmy’ (we did), *zad* ‘zadanie’ (assignment), *poz* ‘Poznań’, *wro* ‘Wrocław’, *swi* ‘Świdnica’, *szcz* ‘Szczecin’, *mchod* ‘Międzychód’.

4.6.13. Initializations

Since initialized phrases are of high frequency, the effort of writing them in full version is managed by means of reducing them to initials. This economy strategy, however,

fails to provide powerful evidence for effort management in text messages as certain initializations are restricted to local use (e.g. *LND*). Some initializations are immediately obvious the issue arises for whom. It might be a pure accident that a person would recognize *TC* (*Take Care*) whereas the highly frequent items *THX* (*thanx*) or *BTW* (*By the Way*) could remain unknown. It is highly likely that certain initializations are popular in specific age, sex or social affiliation groups but not in others, whereas the use of email and internet as well as playing computer games enhances familiarity with acronyms used in SMSes as they overlap (Kul, in press). Examples: *IMHO* (*in my humble opinion*), *ASAP*, *FYI* (*for your information*), *LOL* (*laughing out loud*) and derivatives *LMAO* (*laughing my ass off*), *LMHO* (*laughing my head off*), *TMB* (*text me back*), *PCM* (*please call me*, *p'n'p* 'Piotr i Paweł' (a local chain of supermarkets).

Therefore, the analysis of economy strategies employed both in Polish and English strategies makes the operation of effort management evident. The strategies aimed at managing effort so that the sender put the least work (e.g. omission of punctuation, letter and number homonyms) whereas the receiver's needs were also taken into account in order to avoid communication breakdown (e.g. vowel and middle consonants reduction). As a matter of fact, letter reduction is the prime example of effort management. The sender reduces his/her effort by reducing vowels and at the same time manages to be communicative and carry the content by preserving consonants. The extensive use of economy strategies in SMSes lends more support to the chapter's conclusion that effort management is operational.

Chapter Five

Phonological processes in Natural Phonology

5.1. The aim of the chapter

The present thesis employs the principle of least effort with reference to articulatory effort on one hand, and phonological processes on the other. The former chapters deal with the principle of least effort in articulatory effort specifically, whereas the present and the next chapters are devoted to the principle of least effort in phonological processes. Prior to the discussion of least effort in phonological processes, the use of the term *phonological process* must be clarified with respect to the phonological theory from which the use of the term is derived. The present thesis uses the term *phonological process* exclusively in the way in which Natural Phonology does. First, the chapter describes how phonological processes are understood and used in NP. Section 5.2. explains the general nature and motivation of processes. Then, the chapter deals with the development of processes in child language. Section 5.3. reports on the Stampe's (1969) theory of language acquisition which gave rise to understanding of processes. Next, the chapter discusses the organization of processes. Section 5.4. describes the typology of processes as well as the order in which processes apply. Then, the chapter characterizes the semantic properties of processes. Section 5.5. applies the semiotic principles and parameters to processes. Finally, the chapter overviews Luschützky's (1997) contribution to NP with respect to processes. Section 5.6. presents Luschützky's (1997) typology of processes and his proposal to incorporate articulatory gestures in the concept of a process.

5.2. The general nature and motivation of phonological processes

Phonological process is broadly defined as “a mental operation that applies in speech to substitute, for a class of sounds or sound sequences presenting a specific common difficulty to the speech capacity of the individual, an alternative class, identical but lacking the difficult property” (Stampe 1973 [1979]:1). The concept of substitution as a process is used here metaphorically and therefore should not be treated literally, i.e. that phonological processes consist exclusively in substitution. Since there is substitution for something (e.g. a difficult sound or sound sequence is substituted by an easier one) or substitution for nothing (e.g. a difficult sound is deleted), substitution as a process can mean substitution itself, but it can also mean insertion or deletion as well. Regardless of the process operation (substitution, insertion or deletion), substitution is subject to one underlying intention, i.e. getting rid of the difficulty. For instance, the difficult class of sounds such as voiced stops is substituted with the easier class of voiceless stops by the process of devoicing. On aerodynamic grounds, articulation of voiced stops poses more difficulties than that of voiceless stops when considered in isolation. Therefore, voiced stops are substituted with voiceless stops. In this way, the difficulty embedded in articulation is resolved. These voiced – voiceless substitutions are phonetically motivated since there are also voiceless stops in English. Abundant phonetic evidence suggests that native speakers of English suppress the phonological substitution of obstruent devoicing when in word-final position. A difficult sound sequence may be illustrated with nasal-oral sequence, such as in the word *prince*. This difficult sequence is improved with the aid of epenthesis, which, in fact, is the result of an “out-of-phase” production of the [ns] sequence (*prince* prɪns → prɪnts).

As far as the operational method of substitution is concerned, substitution gets rid of one feature of the sound which is difficult, but never changes more than one feature. In generative theories, it was assumed that one change could involve e.g. three processes at the same time, whereas in NP each change would be ascribed to only one process so that the three changes would involve three subsequent processes. Unlike rule telescoping (Hyman 1975), NP postulates the method of minimal steps in which substitution applies. In a sequence of three processes A, B and C, the application proceeds in the following steps: A → B, B → C. It doesn't proceed from A to C directly (Donegan – Stampe 1979). The operation of substitution is mental, as opposed to

automatic, because there is one intention (to get rid of the difficulty) but many choices to implement that intention (substitution, insertion or deletion). The arguments for the mental character of processes come from various phenomena such as mental speech (where there is a string of sounds in mind without any recourse to the vocal tract) and tongue slips which are planned and systematic (although phonetics distorts the sequence of sounds, they are mentally ordered on the syllable level). Moreover, processes are mentally real: a process can be suspended (it does not have to be applied although there are conditions for its application) or anticipated. Stampe (Stampe 1973 [1979]) makes the following remark on the mental nature of processes: “[a]lthough substitutions are mental in occurrence, they are physical in teleology: their purpose is to maximize the perceptual characteristics of speech and to minimize its articulatory difficulties” (Stampe 1973 [1979]: 9).

Under the above definition by (Stampe 1973 [1979]), phonological processes can be understood a natural reaction to the difficulties posed by speech perception and production. Another definition of *phonological process* is as follows: “a natural reflection of the needs, capacities, and world of its users” (Donegan – Stampe 1979: 127). Processes are *natural reflection* since on one hand they are determined by the vocal tract and by the speech users on the other. *Needs* mean the basic needs to communicate and to express, they are located on the level of psycholinguistics. *Capacities* can be understood in the two ways: either as capacity for speaking (the source is the speech user as the equipment of the vocal tract is designed in order to allow speaking) or as capacity for speech (the source is the speech user in context, his/her identity, style etc., e.g. whether the speech user is a native speaker, a second language learner, an aphasic; whether the speaker addresses his/her speech to a child, a foreigner, another native speaker etc.). *World of its users* means the speech user in context, located on the level of sociolinguistics. Thus, the speech user is subject to his/her vocal tract, but not everything depends on phonetics as they are constraints imposed by language internal system (e.g. sounds inventories, phonotactics) as well as language external factors (age, sex, group membership etc.). The system of constraints is elaborated on and represented by Figure 33.

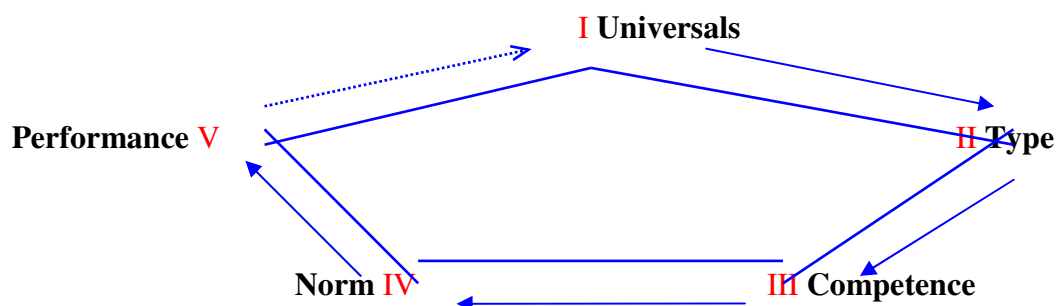


Figure 33. A quintuple: (I) universals (II) type (III) language-specific competence (IV) sociolinguistic norms (V) performance (after Dressler 1985b: 292, Coseriu 1968)

Trask (1996) defined a natural process as “any phonological process which is readily understandable in terms of such factors as the anatomy and physiology of the organs of speech and the acoustic characteristics of speech sounds and which is therefore to be expected in languages. While clearly important, the notion has proved difficult to characterize explicitly” (Trask 1996: 236). Since phonological processes are natural, they are capable of accounting for a wide variety of issues such as “linguistic performance, first and second language acquisition, speech pathologies, casual speech, language games and errors, sound change, silent speech and implicational universals” (Dziubalska-Kořaczyk 2004: 9-13).

As far as the motivation of processes is concerned, they are motivated by the *tension between clarity and ease* (Donegan – Stampe 1979: 130). In other words, their motivation is derived from the constant tension between the contradictory goals of the speaker and the listener. Thus, if the speaker considers a sound or a cluster of sounds difficult to pronounce, he/she overcomes that difficulty by application of the phonological process which “merges a potential phonological opposition into that member of the opposition which least tries the restrictions of the human speech capacity” (Stampe 1969: vii). The nature and motivation of the processes are further explained by their ontological status which is different from that of rules.

Table 8: Processes vs. rules (after Katarzyna Dziubalska-Kořaczyk, p.c.)

Processes	Rules
synchronic phonetic motivation	semantic, grammatical function
innate	learned
apply unconsciously	formed through observation
exceptionless	tolerate exceptions
apply to slips, Pig Latins, foreign words	do not
obligatory or optional	obligatory (conventional, style-independent)

Processes are dynamic, natural and productive since they apply to borrowings, puns, slips of the tongue etc. (e.g. *scotch tape* --> [t^hatʃ sk^jeip] or [k^hatʃ steip]). Rules, on the other hand, have the status of conventions, which means that they lost their phonological productivity and became frozen in the grammar. This means that rules must be learnt from input, must be internalized since they are in the system. They require cognitive effort, they must be learnt and formulated in the course of observation (e.g. *electric* / *electricity*; *pedagogue* / *pedagogy*), unlike processes which are innate and apply subconsciously; in fact, the speaker notices them only if they are not present. The above examples demonstrate the difference in status of processes and rules: rules are prelexical, they participate in word formation (in *electricity* there is a rule /k/→/s/), meanwhile processes are postlexical, apply on an already created word (since according to the rule /k/ → /s/, a context for the process of palatalization arises as /s/ is inbetween /t/). In other words, rules operate in the domain of a morpheme, apply before all processes and feed them. For instance, a rule of vowel lowering i → e (*dream* - *dreamt*) feeds vowel raising before nasal consonant in some dialects of English. Processes have synchronic, phonetic motivation, whereas rules must have semantic, grammatical function. Processes are exceptionless, while rules allow exceptions. Rules are obligatory and must be observed independently of style, e.g. regardless of the casual/emphatic style of speech the rule of umlaut in German in the plural from *Baum* must be applied. Processes can be obligatory or optional since they depend on style.

5.3. Phonological processes in language acquisition

Originally, Stampe (1969 [1979]) proposed NP with a view of addressing the issue of phonological development. He hypothesized that phonological processes are innate. Thus, a child is born with a full set of available processes. The initial set, however, does not resemble a system in the sense of an ordered and organized structure. Rather, the set of processes is being tested out in a creative manner by a child. Consequently, the first productions of infants demonstrate the most extreme processes (e.g. deletion of unstressed syllables as in *banana* → *nana*, cluster simplification, obstruent laxing,

merging of vowels to /ɑ/). The presence of extreme processes can be accounted for by their innateness and by the fact that they all are tested out simultaneously:

[T]here is a fixed universal set of natural processes, and the child about to acquire language 'knows' *all* of these in advance-they are part of his *faculte de langage* on the same footing as other linguistic universals. Moreover, in the first stages of language acquisition he *applies* them all; this accounts for the very restricted phonological repertoire he displays in these early stages (Sommerstein 1977: 233).

The period of extreme processes application lasts throughout infancy. In the course of development, the unordered, unorganized set of processes is subject to change. In the post-babbling period, the child produces the first, nonsemantic words such as *mamama* or *dadada*. These words have systematical properties: they are equally stressed sequences of a lax stop/nasal followed by a vowel. Moreover, these words resemble in systematicity these of adults rather than earlier productions. Even at this stage, the child's phonological representations closely conform to the speech of adults: *doll* is represented as *da*, *dog* is represented as *da* as well. When velars are acquired, *doll* is still represented as *da* but *dog* is represented as *ga* (i.e. the first alveolar stop is assimilated to velar and then the final stop is deleted).

NP hypothesizes that the child selects and applies these processes which he/she encounters in the speech of adults: "[t]he phonological system of a language is largely a residue of an innate system of phonological processes revised in a certain way by linguistic experience" (Stampe 1969 [1979]: vii). For instance, a German child is at liberty to produce devoiced obstruents. An English child, however, must suppress the natural process of obstruent devoicing because the phonology of the English language does not allow it. This means that the child's own phonological system is constantly subject to revision towards the adult's phonological system. The revision consists in restriction and inhibition of processes which do not conform to the adult system and last throughout the period of language acquisition: "[t]he child's task in acquiring adult pronunciation is to revise all aspects of the system which separate his pronunciation from the standard. If he succeeds fully, the resultant system must be equivalent to that of standard speakers" (Stampe 1969 [1979]: x). The goal of revision may be achieved via implementation of three mechanisms: suppression, limitation and ordering. First, the differences in the child's system are compared to the adult's system and subsequently limited to specific sounds or sound sequences (limitation). Next, the substitutions which appear in a random or an unordered way become ordered in the child's system. A sound

represented by a variety of sounds in the child's language receives a more systematic representation (ordering). Finally, the processes in the child's language which are not consistent with the adult, system become suppressed (suppression). Each of these mechanisms can be illustrated with examples from child language. Suppression: ki: → (suppression of prevocalic tension) → kɪi (suppression of postsyllabic desyllabification) → kɪri (suppression of flap deletion) → kɪti (suppression of flapping). Limitation: voicing obstruents before voiced segments (*papa* represented as *baba*), then voicing obstruents inbetween voiced segments (*papa* represented as *paba*) and elimination of the process of voicing between voicing segments (*papa* represented as *papa*). Ordering: there are two processes: process A flapping of /n/, flap deletion and desyllabification (*changing* represented as *kæɪŋ* and *candy* represented as *kæɪ̃*) and process B /nd/ → /nn/ → /n/. If the process B is applied before the process A, *candy* is represented as *kæɪ̃*. Later, the process A is applied before the process B, *candy* is represented as *kæ̃nɪ*.

The acquisition mechanisms of suppression, limitation and ordering are implemented also in the adult life as they resolve the contradictions between processes. These contradictions arise from conflicting phonetic restrictions. For instance, obstruent devoicing, on the one hand, is a context-free process (due to occlusion impeding the airflow required for voicing), whereas obstruent voicing is context-sensitive on the other hand (by assimilation in an intervocalic position). Since no obstruent can be voiced and voiceless at the same time, adult phonology selects and implements one of the three mechanisms by which the child adjusts his/her phonology: suppression (the English language suppresses the process of obstruent devoicing), limitation (i.e. partial suppression of the process in terms of segments or contexts – e.g. tense obstruents or intervocalic context in English) or ordered application (e.g. devoicing in word-final position and the later application of voicing intervocalically).

Not only is NP capable of accounting for phonetic representations of children and adults, but also for phonemic inventories. With the help of the innateness and revision prediction, the regularities reported in child language can be accounted for without recourse to implicational laws. This means that the results of implicational laws derive from the innate system itself. The implicational laws hold, e.g. that affricates imply spirants, which, in turn, imply stops. The two context-free processes also account for these laws: obstruents become stops (to be more specific, affricates become stops)

and affricates become spirants. Thus, the acquisition order justifies the order implied by implicational laws. Furthermore, innateness and revision prediction account for language change: “[a] phonetic change occurs when the child fails to suppress some innate process which does not apply in the standard language” (Stampe 1969 [1979]: xvii). In this way, NP departs from the view that the child possesses a phonological system of his/her own and in its own rights. In NP, no special status is assigned to child phonology. The confusion stems from the treatment of the early, extreme productions. These productions are rejected by the standard. In some dialects of English, however, these allegedly deviated productions are accepted by the standard. A child fails to suppress the process of obstruent devoicing. This results in a change which admits devoicing. The contrast is retained and marked by the length of the preceding vowel. A vowel is longer before a voiced sound and shorter before a voiceless one.

5.4. The organization of phonological processes

The phonological behavior of children and adults displays seemingly contradictory processes. Some of adult native speakers of English denasalize foreign words containing nasalized vowels (French *maman* [mɑmɑ̃n]) → [mɑmɑ]) and nasalize the vowels followed by nasals in English words (can’t [kæ̃nt]). The child first nasalizes vowels (*lion* represented as [nã̃n]), then denasalizes the vowel in the same word ([nã̃n] → [nɑn]).³⁷ Vowel denasalization eliminates the difficulty of the nasal gesture, whereas nasalization magnifies the difficulty and spreads the nasal feature onto the neighboring vowel. These two processes contradict each other only seemingly. Vowel denasalization is a context-free process (paradigmatic) which generally serves to avoid another feature to a vowel since nasal vowels are more marked than the oral ones. Vowel nasalization, on the other hand, is a context-sensitive process and it is natural (unmarked) to nasalize a vowel in a nasal context. The latter process is context-sensitive as the sequence condition (a vowel followed by a nasal) must be met. Thus, these contrary phonetic conditions give rise to the two types of processes: context-free (fortition) and context-

³⁷ The data come from a child named Joan Velten.

sensitive (lenition). These types apply in the contrary speech styles: fortitions in slow, formal speech and lenitions in fast, informal speech. Fortition governs underlying representations (phonemes) and acts as a measure to maintain sound distinctivity (denasalization bars nasal vowels from the English language). Lenition governs surface representations (allophones), thus it is an allophonic process (Gurevich 2004). English has no nasal vowels but any vowel can have its nasalized allophones. Lenition “gives rise to sounds eliminated by a prior, more general process in the system” (Stampe 1973 [1979]: 27). Therefore, the organization of phonological processes proceeds under the lenition/fortition typology. This typology introduces the issue of process ordering.

The order in which the processes apply is random and sequential. This means that one process applies to the output of another and excludes simultaneous application. For instance, in the word *button* [bʌʔn̩] the two processes: deglottalization and nasalization cannot apply at the same time. First, the glottal stop is deleted ([bʌn̩]) and nasalization occurs on the output of deglottalization ([bʌ̃n̩]), otherwise, the intervening glottal stop bars the nasal feature from spreading onto the vowel. In child language, the sequential order is even more evident. The child produced the representations of the word *lamb* in the following order: [jæm] (delateralization) → [ʒæm] (spirantization) → [zæm] (depalatalization) → [zæb] (other processes): “if the processes had applied simultaneously, she [the child] would have ended up having to pronounce two-thirds of the sounds the processes are supposed to eliminate” (Stampe 1973 [1979]: 63). Thus, the substitutions appear and disappear separately. The natural, sequential order is more logical than explaining the subsequent representations as a rule under which the voiced oral apical continuants /l, j, ʒ / change into /z/. Rather, the change of /l/ into /z/ entails the intermediate steps in the form of /j/ and /ʒ/ substitutions. This sequential order, postulated by Stampe (1973 [1979]) is non-linear: “processes interact with each other [...] apply and re-apply whenever the configurations they eliminate arise (Stampe 1973 [1979]: 60). In other words, the processes apply when there is context for them. If the context which has been eliminated by a previous process is restored by another process, the process reapplies. This means that the intrinsic order of phonological processes is also iterative (reapplication possibility) and natural (no context, no process). The

iterative, sequential and natural order derives from the functions particular process types serve.

Phonological processes are organized into types (lenitions/fortitions) and ordered (non-linearity). Ordering has consequences for the two types and its sequential property means that the output of one process constitutes the input to another process. Since processes are simultaneous on one hand, but sequential on the other, it means that many of them can be applied (simultaneous) but not at the same time (sequential), thus the application of many processes proceeds in steps. Only after a process number one has been applied, a context for a process number two is created and only then the process number two can be applied (otherwise the context is not there). Thus, the process number two feeds on the result of the process number one and this relation is called feeding. Feeding arises when the output of the process A creates input for the process B. A feeds B, i.e. the process B applies to the output of the process A. Derivation of *divinity* ([dəvɪnɪtɪ]), exemplifies feeding: process A – flap deletion ([dəvɪnɪ]), process B – nasalization ([dəvɪ̃ɪ]). Other examples of feeding: *batted* [bætəd], process A – vowel epenthesis [bætɪd], process B – tapping [bæɾɪd]; *meant* [ment], process A – nasal deletion [met], process B – vowel nasalization [mẽt]. Feeding is the case of natural order. The derivation terminates after all available processes have applied. Feeding allows reapplication of a process, if a process number one creates the context for a process number two, which can create a context for a process number three etc, and then a process number four may be the same as process number one. Therefore, counterfeeding is a constraint on iteration since reapplication of a process may result in too big a dissimilarity of the initial form in relation to the final one: *pat it* [pæɾɪt] is flapped when applied to the basic representation. Flapping, however, does not apply to derived representations such as *plant it* [plæ̃ɾɪt] where nasal elision counterfeeds flapping. This constraint has phonetic motivation and makes the basic representation closer to the phonological intention.

Processes apply in an iterative and sequential order (feeding), but their own, intrinsic order is random. Bleeding, unlike feeding, stops a process from applying. Bleeding arises when the process A eliminates the potential input to process B since A removes the context for B. The output of A blocks application of B: e.g. A - vowel

raising, B - consonant deletion, *betting* [bɛrɪŋ] does not become [bɪɪŋ], thus vowel raising bleeds consonant deletion. Another example is *three* [θri] → [θɹi] (B), but: [θəri] (A) and not *[θəɹi] since A (vowel epenthesis) bleeds B (flapping). More important, feeding and bleeding establish the relationship between lenitions and fortitions: fortitions have priority over lenitions and are ordered (apply) before lenitions. Fortition may feed lenition: *warmth* [wɔ̃mθ] → [wɔ̃mpθ] → [wɔ̃pθ] (consonant /p/ epenthesis fed assimilation, this was possible as the consonants /p/ and /m/ are homorganic) or bleed them: *batted* [bæt + d], fortition: vowel epenthesis ([bætəd]), lenition: voice assimilation like in *kissed* [kɪs + d] → [kɪst] but the form [bætət] is not possible any more and blocked by the epenthesis, epenthesis bleeds assimilation like in *Timothy* [timiθi] → [timθi] → *[timpθi] (consonant /p/ epenthesis could not feed assimilation, this was impossible as the consonants /p/ and /θ/ are not homorganic). If processes apply simultaneously, counterbleeding results. Regressive assimilation is not bled by nasal elision: *can't* [kæ̃nt] → [kæ̃t] but *can't* will never become [kæt] due to the fact that the superficially nasalized vowel cannot be denasalized, so the process does not really apply as vowels are nasalized before nasal consonant in English.

5.5. The semiotic principles and parameters in phonological processes

Natural Phonology views phonological processes as “a natural reflection of the needs, capacities, and world of its users” (Donegan – Stampe 1979: 127). The concept of naturalness, however, seems to be problematic and vague if is considered without context. In order to make the concept more meaningful and give NP a proper epistemological frame, Dressler (1985a, 1985b) adopted semiotic as a metatheory of NP. Thus, semiotics serves as an explanatory model of phonology in NP. There are numerous reasons for which the science of signs has become the NP metatheory. Firstly, semiotics addresses naturalness in a gradual way, treating naturalness as the other end of the natural ↔ conventional spectrum. Secondly, semiotics is a theory of signs and

language is commonly defined as a system of conventional signs, primary spoken, secondary written (Gimson 1995). Thus, semiotics helps to explain the functions of linguistic signs. If, for instance, processes and rules are treated as linguistic signs (cf. 5.2.), rules are conventional signs, least natural as their operation is not immediately evident (such as in the plural form *child* from *children*), whereas processes are natural, transparent signs as the speaker is aware why s/he applied them. Thirdly, semiotics is functional and so is language (e.g. the function of language is communication). It must be clarified that it is not general semiotics which serves as the NP metatheory, but rather that of Peirce (1991). He employed the idea originally proposed by de Saussure (1915 [1983]) that a sign has two aspects: the signifiant (i.e. the representation of a sign or its form, the signifying) and the signifié (i.e. the object which is represented or its substance, the signified). Peirce (1991) complemented these two fundamental aspects of sign with two more aspects: that of an interpreter and that of an interpretant. An interpreter is the person who interprets a sign, otherwise the sign becomes meaningless without the signifying agent. Moreover, an interpreter cannot function without the respect or capacity (an interpretant) in which a sign performs its functions for an interpreter. In other words, an interpreter ascribes significant to significates on the basis of an interpretant, i.e. an idea (Dressler 1985b). Peircean semiotics serves as a suitable candidate as the basis of naturalness since these two theories overlap in many respects. For instance, Peircean semiotic (1991) is processual and has numerous direct correspondences to linguistic concepts. Moreover, Peircean semiotics (1991) bridges the gap between form and substance due to connecting chains of signs (form) with phonetics (substance) on various levels of the linguistic structure.³⁸ Therefore, adoption of semiotics, in particular, that of Peirce (1991) makes all claims against naturalness unwarranted (Dressler 1985a, 1985b).

Therefore, the structure of NP derives from semiotic principles and parameters. They, in turn, are applicable to phonological processes. Semiotic makes use of the figure-and-ground principle which can represent different objects, according to what an interpreting person considers to be the figure and what The semiotic principle of figure-and-ground considers to be the ground. If the principle is visualized (Figure 34), an interpreting person may focus on the foreground and see a vase (the figure) or, an

³⁸ The concept *chain of signs* denotes the multiconnectivity of signs: one sign stands for another, which, in turn, stands for a sign for a sign for a sign and so on.

interpreting person may focus on the background and see two faces (the ground). The interpretation depends on what an interpreting person considers to be the ground or the figure.



Figure 34. A visualization of the semiotic principle of figure-and-ground (website Figure/ground at <http://www.psychologie.tu-dresden.de/>).

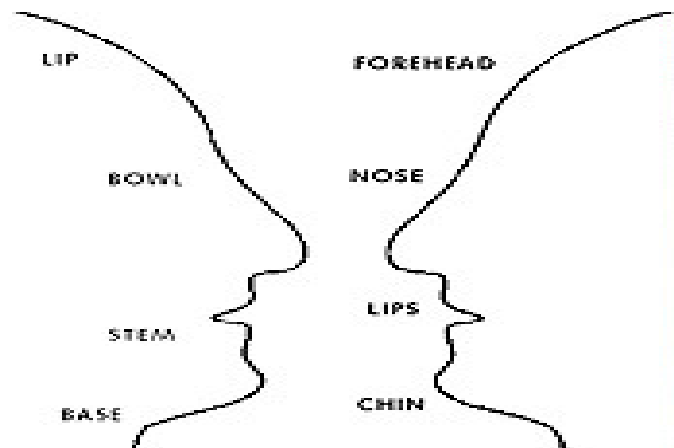


Figure 35. Labelled figure-and-ground effect (website Figure/ground at <http://www.psychologie.tu-dresden.de/>).



Figure 36. A picture representing the semiotic principle of figure-and-ground (website Figure/ground at <http://www.psychologie.tu-dresden.de/>).



Figure 37. A vase representing the semiotic principle of figure-and-ground (website Figure/ground at <http://www.psychologie.tu-dresden.de/>).

The semiotic principle of figure-and-ground has been applied to lenitions and fortitions. These two sets of processes can be represented as figures or ground according to the functions they perform (for a more detailed discussion, cf. Chapter Six, subsection 6.3.2). Fortitions serve the foregrounding of figures, whereas lenitions serve the backgrounding of grounds. Predictions concerning the distribution of foregrounding and backgrounding processes can be made on the basis of the treatment of phonemes in the sense of Baudoin de Courtenay (1877 [1972]) who treated a phoneme as a sound intention (*der Lautabsicht*). Sound intentions must be adapted to the needs of production and perception and, as a result, sound intentions must overcome the difficulties posed by the two needs. If one considers the formal speech as the reflection of the needs of perception and casual speech as the reflection of the needs of production, one can predict the distribution of processes. Thus, formal speech is the domain of foregrounding processes (fortitions), whereas casual speech is the domain of the backgrounding ones (lenitions): “formal speech situation can be characterized as those where the speaker subordinates his speech strategies to the exigencies of the hearer(s) (...) casual speech situations can be characterized as those where the speaker is less concerned about understandability of his production” (Dressler 1985: 33). Furthermore, fortitions apply prior to lenitions. This order derives from the fact that perception of phonological intention presupposes the adaptation of the speech tract. From the diachronic perspective, both foregrounding and backgrounding processes apply:

foregrounding processes as baby-talk helping to acquire sound intentions, backgrounding processes generalize from casual to obligatory processes. Fortitions, however, are morphologized into rules like the English Great Vowel Shift or are lost. The above observation highlights the distinction between processes (e.g. devoicing) and rules (e.g. the umlaut rules in contemporary Germanic languages) which is equivalent of to that of Donegan – Stampe (1979). Dressler (1985a) introduced graduality into this distinction in terms of a scale from prototypical phonological process to prototypical morphological rules. Predictions concerning prototypical properties of processes and rules can be made on the basis of the two main functions of segmental phonology: ease of articulation and clarity of perception.

The next set of predictions concerns necessary or prototypical properties of processes vs. rules. I want to illustrate this with Polish velar palatalisations [...] The only process is velar surface palatalization (SP1) which turns all velars /k, g, x/ (written <k, g, ch>) into [k', g', x'] before /i/. There is also a MPR of surface palatalisation (SP2) which turns only /k, g/ into [k', g'] before *e*-initial suffixes. For example, the instrumental case form of the name *Maks Plank* is either *Maks-em Plank* or *Maks Plank-iem* with a final syllable [k'em]. There is a phonemic contrast between [k', g'] and /k, g/ before /e/ but never before /i/. Then we have an AMR of alveolar formation (AF) which turns /k, g, x/ into *c* [ts], *dz* [ʒ], *sz* [s] [...] Finally, I will use the MPR of spirantisation (SPIR), which turns *dź* [ʒ] (derived by PF from /g/) into *ź* [ʒ] after vowels/sonorants and before certain suffixes (Dressler 1985a: 36-37).

Thus, in Polish there is one process and three rules: (1) a process of velar palatalization /k, g, x/ → [k', g', x'] /_ /i/ (2) a MPR (MorPhonological Rule) /k, g/ → [k', g'] /_ +e... *Plank* → *Plankiem* (3) a AMR (Allomorphic Morphological Rule) e.g. *Polak* → *Polac+y* (4) a MPR (spirantization) *Bóg* → *Boż+e*. Velar palatalization generalizes to casual speech (e.g. *wróg Ireny*) and acts as a constraint on pronounceability. Velar palatalization is a phonetically motivated process, it is exceptionless and productive, whereas rules (2), (3) and (4) can be productive (neologisms, loan-words) but, unlike the process (1), do not apply to slips of the tongue. This example allows formulating predictions concerning natural ordering and establishes the primacy of morphology over phonology. This primacy has consequences for the order of processes and rules: rules are prior to processes. Using the framework of semiotics, processes and rules can be represented as signs. The signantia (allophone) of a sign (phoneme) ought to be distinguishable, which

entails maximum differentiation and minimization of fusion in formal speech. Casual speech allows fusion and deletion. Distinctivity of signantia, in turn, entails its concreteness and effectiveness in terms of signs:

It must neither be too big nor too small, for otherwise the sign might not be easily produced or perceived. This principle allows the predictions that (1) ultrashort sounds will be rare and easily lost; that (2) vowels will be more frequent than diphthongs or triphthongs, and stops and fricatives will be more frequent than affricates (Dressler 1985: 36).

Dressler (1985a) systematized the issue of process universality concluding that the motivation of phonological processes is universal, whereas processes as such are language – specific. For instance, both Polish and German apply the process of word-final obstruent devoicing which does not apply in English. Therefore, a German or a Polish speaker is at liberty to devoice word final obstruents which is a natural, universally motivated process. On the other hand, a native speaker of English is required to suppress this process in his native language: “the universality of processes does not mean that they apply in all languages - only that they are motivated in all speakers” (Donegan 2002: 64). On the basis of universality, Dressler (1985a) introduces a distinction between process types which are universal, natural processes (e.g. devoicing) and processes which are language-specific (e.g. syllable-final devoicing in German, word-final devoicing in Russian or Polish). In other words, process type is a general tendency to devoice, whereas a process is its concrete, specific realization in a given language which, as a language-specific phenomenon, has constraints on distribution (e.g. final obstruent devoicing in Slavic languages). He motivated the need for the distinction between a process and a process type in the following way: “phonemes and their phonetic representations are the outputs of phonological processes which are the remnants of universal natural processes inhibited in language acquisition” (Dressler 1985a: 30). Universal processes types possess inherent hierarchies. An example of hierarchies is reduction of nasal assimilation hierarchy to three thresholds: total suppression, assimilation of /n/ to following plosives and most general, assimilation of /n/ to all obstruents in its vicinity. In German, the first type occurs in formal speech, whereas the second type can be attested in casual speech. The most general process type was encountered in Wernicke’s aphasics who produced [gʏali] for *gemalt*.

The semiotic parameters apply to phonological processes and are the basis of their frequency predictions. A sign can be represented as an icon, an index or a symbol. These three types of signs can be arranged on a scale from the least natural (symbol) to the most natural (icon): symbol → index → icon.³⁹ Figure 38 illustrates the concept of male and female in terms of the symbol - index - icon scale.

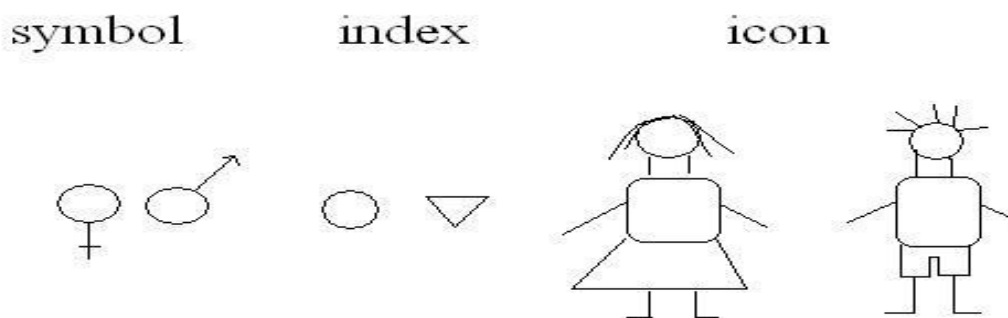


Figure 38. The concept of male and female in terms of the symbol - index - icon scale.

Iconicity is the property of a sign which ensures a sign's transparency and naturalness. It means the smallest the distance between input and output. Iconic relations are similar or analogous. The parameter of iconicity allows making predictions concerning frequency. Iconic relations should be most frequent since icon is the most natural type of sign. Moreover, iconic relations are simpler which is also reflected by frequency: simpler phenomena are more frequent. The above correlation between simplicity and frequency accounts for higher frequency of processes in comparison with rules. Processes have synchronic, phonetic motivation which makes them simpler whereas rules must be internalized with time. Moreover, processes make minimal changes so that the distance between the input and output is insignificant and easy to perceive immediately. Beside, the system of a language treats rules secondary to processes in terms of importance for languages and language types. Rules do not exist in isolating languages, where processes alone suffice for the system. Another consequence of iconicity is that intrinsic allophonic processes are far more frequent than extrinsic allophonic process and even more frequent than phonemic processes. This is due to the distance between input and output which in the case of intrinsic allophonic

³⁹ Icons, indexes and symbols, which were identified and named by Peirce (1991), are types of signs. They capture the abstract criteria of sign systems such as the links between signs and the links within signs. More specifically, they comprise a signaling system which links the object of perception, the perceiver and the method by which she/he perceives the object. A symbol is the cognitive reception of arbitrary signs, the mental effect which depends on the presence of an *interpretant*. An index is a sign which does not require the presence of an interpretant but does require the presence of its object or form. An icon is a sign which would possess the characteristics which render it significant even though its object had no existence.

processes is relatively big. Iconicity for input – output relations is defined by the number of distinctive features. This also holds for natural classes of sounds which by definition have the common denominator of distinctive features: classes of sounds are input for the most iconic processes (e.g. the class of velars is the input for the frequent process of velar palatalization in Polish). Iconicity in terms of simplicity has consequences for aphasia. The simpler a structure is, the less disturbed in aphasia it becomes. Evidence from Polish demonstrates that an intrinsic allophonic process (*kot* ‘cat’ sg. → *koty* ‘cats’ pl.) was least disturbed, whereas the allomorphic morphological rule (e.g. *Polak* ‘Pole’ sg. → *Polacy* ‘Poles’ pl.) was most disturbed.

Diagram represents an analogy between input (phoneme) and output (allophones). Diagrammaticity is the similarity of input and output, however, unlike iconicity, does not treat similarity in terms of naturalness but in terms of similarity of the forms before (input) and after (output) a process application. For instance, the forms before and after voicing are similar whereas the forms before and after metathesis are not. Thus, diagrammaticity is the similarity parameter for processes which predicts that the most diagrammatic processes are the most frequent, deletion and fusion are less frequent (in formal speech only) and metathesis is the rarest (Dressler 1985a: 41):

diagrammatic /A B/ /A B/
 | | | |
 [a b] [a' b']

e.g. voicing

non-diagrammatic /A B C/ /A B C/
 | | | |
 [a Ø c] [a c]

e.g. deletion and fusion

anti-diagrammatic /A B/
 X
 [b a]

e.g. metathesis

Indexicality means a direct connection between input and output. It stipulates a certain relation between input and output in the sense that input leaves a trace on the output (e.g. nasal/input leaves a trace in the form of nasalization/output). It also favours those relations where the distance between input and output is relatively big (optimal). The bigger the distance, the better it becomes to perceive since a contrast, unlike similarity, is optimal for perception. Indexicality is characteristic of those neutralizations which leave traces (e.g. anaphoric pronoun leaving trace of a coreferential noun). Context-sensitive processes (lenitions) are indexical and frequent because the context leaves a trace (e.g. nasalization is a trace of a nasal sound). Context-free processes (fortitions), on the other hand, are non-indexical and rare. Intrinsic allophonic processes are also non-indexical (or least indexical) due to the fact that the small distance between input and output impedes perception. Intrinsic allophonic processes, however, are more frequent than extrinsic allophonic processes and even more frequent than phonemic processes. This frequency order is accounted for by iconicity which is prior to indexicality as it is a more natural and transparent sign type. In similar vein, the primacy of morphology over phonology explains why morphological indexicality is more important than the phonological one. Indexicality alone is capable of accounting for a change of a process into a rule. Diachronically, a rule never becomes a process again because the optimal distance between input and output is unlikely to minimize, thus, efficiency of a rule would decrease. Indexicality parameter is also important for obligatory processes/rules as a trace is left (a phoneme to which a process assimilates is indicated), whereas it is not necessarily so for optional processes/rules (a trace may or may not be left).

Biuniqueness is a preference and is better than uniqueness or non-uniqueness. As a matter of fact, iconicity favors biuniqueness so that biuniqueness and iconicity predictions converge. Thus, processes are most frequent (biunique), whereas rules are less frequent (no-unique). Moreover, phonological biunique and unique relations are productive and general (e.g. neutralizations).

strong biuniqueness (Dressler 1985a: 44-45):

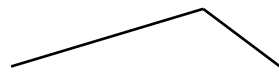
/A/ /B/

| |

[a] [b]

as in unnoticeable intrinsic allophony

weak biuniqueness: /A/

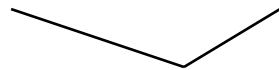


[a] in condition C [b] in condition D

as in extrinsic allophony

OR

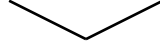
/A/ in condition C /B/ in condition D



[a]

as in certain cases of phonemic overlap

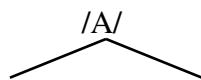
uniqueness: /A/ /B/



[a]

as in neutralization

OR

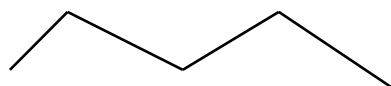


[a]

[b]

as in free variation

non-uniqueness: /A/ /B/



[a]

[b]

[c]

here /A/ → [b] and /B/ → [b] are non-unique.

The above diagrams illustrate biuniqueness, uniqueness and non-uniqueness. In uniqueness there is A and B, and if B derives from A, there is a twodirectionality (the relation works in two directions) as there is a relation between A and B (A is input of B) and there is a relation between B and A (B is output of A). In biuniqueness, B derives from A but A does not derive from B so that the relation is onedirectional. For instance, biuniqueness holds between phonemes (A) and allophones (B). The relation works only

in one direction as B ([t^h]) comes from A (/t/) but A does not come from B as /t/ has many other allophones: /t/ → [t] *cat*, [t^h] *tap*, [t̚] *Hatchepsut*, [t^ɰ] *cotton*, [tʰ] *little*, [t̪] *eighth*, [t̚] *try*, [ɾ] *bottom*, [ɹ] *dirty*. Non-uniqueness represents too many possibilities and relations to be established. For instance, glottal stop can be an allophone of every phoneme in English. Another good example comes from Polish, where /wat/ (plural from *wata* “cotton”) and /wad/ (plural from *wada* “fault”) results in two allophones ([t]₁ and [t]₂) and two phonemes (/t/ and /d/).

5.6. Luschützky (1997)

Luschützky (1997) is not directly affiliated with NP. Nevertheless, his approach to phonological processes partly employs the treatment of processes as developed by Dressler (1985a, 1985b) and Stampe (1973 [1979]). Luschützky (1997) shares the view that processes are natural because they derive from the properties of human production and perception. The use of the adjective *phonological* with reference to processes, unlike for Dressler (1985a, 1985b) and Stampe (1973 [1979]), raises certain objections for Luschützky (1997) since the clear-cut division between phonetics and phonology is artificial and can be compared to another artificial division between physics and chemistry. Rather, processes should be viewed as an interaction between phonological intentions and phonetic behavior. Luschützky (1997) also criticizes the claim that phonological processes are phonetically motivated which has been revised by Ronneberger-Sibold (1984). She proposed the structural motivation of processes, arguing that processes are driven by economic preferences, reducing the segment inventory and simplifying the system’s structure. This proposal is criticized by Luschützky (1997) as not explanatory adequate since phonological processes must interact with morphology (derivation), thus fail to simplify inventories on their own. Moreover, there is no metaprocess or a global process which could have economizing function. Luschützky (1997) agrees on the natural character of processes, their phonetic motivation and their typology. The distinction between lenitions and fortitions, for instance, is applicable to both prosodic and segmental domains. Prosodic domain serves

as the interaction platform between the phonemic arrangement and the intonational/rhythmic patterning:

Table 9. Process types in prosodic and segmental domains
(after Luschützky 1997: 68)

	Prosodic	Segmental
Fortitions	Vowel lengthening	Diphthongization
Lenitions	Vowel shortening	Monophthongization

In the NP literature, lenitions and fortitions are also referred to as syntagmatic (operating on sound sequences), context-sensitive, centripetal (lenitions) and paradigmatic (operating on individual sounds), context-free, centrifugal (fortitions). The issue of paradigmatic dissimilation, however, seems to demonstrate that the notion *centrifugal* with reference to fortitions is misleading. Paradigmatic dissimilation guarantees clear perception. The nature of perceptual space in the case of consonants does not suffice to differentiate between the center and the peripheries. To take the analogy further, paradigmatic processes which aspirate fortis stops in order to differentiate them from the lenis ones would introduce no such difference since the perceptual center is not a distinct space. Moreover, the difference would be only in terms of relative opposite positions (fortis - lenis). Thus, the differentiation fails to account for absolute positions in analogical cosmos of perception. In similar vein, the classification of the vowel coloring process as fortition can be called into question (Donegan 1985).

As far as the theory behind process typology is concerned, Luschützky (1997) departed from the markedness theory since the markedness parameter is binary. As such, it fails to account for the dynamic aspect of processes: “[e]ine tiefeschürfende Analyse des Markiertheitsbegriffs ist daher für die Belang der Prozesstypologie nicht erforderlich” “[Any profound analysis of the markedness concept is not necessary for the issue of process typology] (Luschützky 1997: 46).⁴⁰ He also questioned the explanatory power of the sonority hierarchy and consonantal strength. These two parameters do not explain process typology in a reasonable way, nor explain the diachronic aspect of sound growth. Thus, in the absence of a good theory, an alternative approach to process typology which goes beyond the strength of sounds should be developed: “[a]lle

⁴⁰ All the translations from German sources are mine, MK.

bisherigen Versuche, die Menge möglicher phonologischer Prozesse durch verbindliche Sonoritäts- und Stärkehierarchien zu limitieren, müssen als Fehlschläge bezeichnet werden“ [All the attempts which have been undertaken so far, and which attempt to limit the possible phonological processes via sonority and strength hierarchies have failed] (Luschützky 1997: 148). In order to provide a more suitable explanation of process typology, he employed the notion of *articulatory gestures* (Browman – Goldstein 1986, 1992) which derives from the motor theory of speech (Lieberman – Mattingly 1985).

Under Luschützky's (1997) approach, gestures are interpreted as phonetic implementation rather than the results of jaw movements: "die Bewegungen des Unterkiefers mit der akustischen und perzeptionellen Seite des sprachlichen Signals nur sehr indirekt verbunden sind und dass auch diese lose Verbindungen keine mandatorische ist" [the movements of the lower jaw bears merely indirect connections with the acoustic and perceptual aspect of human speech and even these loose connections are not obligatory] (Luschützky 1997: 165). Therefore, he proposed the following constriction parameters in accounting for the nature of articulatory gestures: a) the degree of constriction (opening, approximation, friction, occlusion), b) the completeness of constriction (local: central – lateral, temporal: occlusive – vibrant – ballistic) and c) gestural condensation (contour of segments: affricates, diphthongs, complex articulations: labial – velar). These parameters display a direct connection with articulatory movements. For instance, the difference between completeness and degree of stricture has consequences for delateralization processes of coronals. Luschützky (1997) employed the classic quartet of gestures: initiation (laryngeal, pulmonic, velaric), phonatory (voicing), nasalization (oral, nasal) and articulatory ones (opening, approximation and closing).⁴¹ There is a hierarchy of gestures in terms of production: initiation gestures are accountable for by the difference in air pressure, whereas phonation handles the strength of sounds as well as additional contrast. With the aid of articulatory gestures, a specific shape of sounds is assumed.

Having incorporated the notion of gestures into process typology, Luschützky (1997) proposed the following process types:

⁴¹ In the phonetic literature, nasal gestures are frequently subsumed under the articulatory ones. This is not correct since nasal gestures constitute a separate category. The autonomous status of nasal gesture is supported by distinct innervation of the palatoglossus muscles (Luschützky 1997).

- a) Intrasegmental - concerning gestural composition of a single phoneme. The intrasegmental processes are context-free and comply with the articulatory gesture of the phoneme they substitute.
- b) Intersegmental - concerning gestural coordination of the adjacent phonemes. They are motivated by the locality of segments and encompass such phenomena as contact assimilation, contact dissimilation and contact metathesis, specifically epenthesis and coalescence.
- c) Transsegmental - concerning gestural interaction of non-adjacent phonemes. They affect groups of sounds consisting of two or more members which are not in the vicinity of one another. These processes include distant assimilation, distant dissimilation and distant metathesis, in particular the instances of haplology.

These processes perform the functions of lenition or fortition. The choice of a function is determined by the temporal parameter (i.e. how gestures are organized in time) and attention level. Temporal organization of gestures is governed by the factors of duration, inherent carrying movement and the intrinsic profile of movement. These factors are particularly significant in the case of intersegmental and transsegmental processes. Assimilation of voice is a powerful piece of evidence of temporal coordination driven by transsegmental process. The above typology complements the paradigmatic-syntagmatic distinction in terms of segment distance, extending the notion of syntagmatic processes. Moreover, the intra-, inter- and transsegmental distinction is drawn in order to demonstrate that gestures exceed the idealized limits of linguistic units (here segments):

Das Beispiel von Epenthesen [...] zeigt sehr deutlich dass man segmentale Prozesse zwar so nennt, weil sie Einfluss auf die substantielle Gestaltung von Segmenten haben und teilweise hieraus ihre Motivation beziehen, dass jedoch andererseits der zugrundeliegende Mechanismus in der zeitlichen und dynamischen Koordination relativ autonomer Komponenten des Produktionsaggregats besteht, die sich nicht weit über die idealisierten Grenzen abstrahierter Einheiten hinaus erstrecken (Luschützky 1997: 69).

[The example of epenthesis demonstrates clearly that segmental processes were named the way they were named since the impact of the shape of segments involved has also involved its motivation. It also demonstrates that although on one hand the mechanism of underlying temporal and dynamic coordination constitutes the relative autonomy of the product components, on the other hand the mechanism hardly goes beyond the domain of abstract entities].

Chapter Six

Phonological processes and the principle of least effort

6.1. The aim of the chapter

The aim of the chapter is to revise the current typology of phonological processes. The revision is based on the principle of least effort and seeks to develop alternative lenition/fortition definitions and typology. First, the chapter deals with the notion of typology. Section 6.2. outlines, among others, the difference between a typology and a classification. Next, the chapter discusses various approaches to lenition/fortition. Section 6.3. presents the current approaches to process typology: the traditional one, the NP and the OT approaches and concludes that no coherent definition of lenition/fortition exists. Then, the chapter proposes a definition of lenition/fortition in terms of three criteria. Section 6.4. discusses the three criteria of lenition and presents definitions of lenition/fortition. Next, the chapter revises the typology of phonological processes under effort management. Section 6.5. lists the concrete, specific lenition/fortition processes and proposes a new typology. Finally, the chapter discusses teleology. Section 6.6. establishes the primacy of the teleological explanation over the causal one and adopts teleology into effort management.

6.2. Typology

Human thinking is cognitive in the sense that it strives to explain various objects and phenomena of the surrounding world. In explaining, it also tends to classify objects and phenomena, both material (e.g. plants, animals) and non-material (e.g. zodiac signs, personality types). Classification proceeds according to parameters which take into consideration and evaluate the visual, acoustic, functional, morphological etc. features of the classified objects. Certain features, such as functional and morphological, are fuzzy due to an overlap. For instance, the shape of an object reflects both its morphological and functional features. Nevertheless, the features serve as parameters which classify an object into a type. Classification means grouping objects on the basis of the lack or presence of a feature. Typology, unlike classification, means grouping objects on the basis of a theoretical background (Luschützky 2005). For instance, diachronic comparative linguistics offers a typology of languages on the basis of the cognates and language families' theory. The theory stipulates the existence of a common ancestor (e.g. Latin for all Romance languages). In contrast, synchronic comparative linguistics represents a classification since it orders languages into classes (called conventionally *types*) according to their properties (e.g. there are isolating, inflecting or incorporating groups which classify Latin and Polish as inflectional languages). In comparison with diachronic comparative linguistics which uses the common ancestor theory, the synchronic comparative one classifies languages solely on the basis of the way a language grammar reflects the relations between the objects (languages with developed inflection systems are classified as inflectional). Thus, typology is a selective classification since it involves parameterization of features that have a certain cross-componential validity (Moravcsik 1979).

The usefulness of a typology is determined by the number of types which a typology incorporates. The fewer types are postulated, the fewer entities fit them. By analogy, the more types are established, the easier it becomes to typologize entities.⁴² Accordingly, there exists a metatypology, i.e. the typology of typologies. A typology can be hierarchical, dualistic or optimal. A hierarchical typology is based on a rank a certain phenomenon occupies in a hierarchy. An optimal typology provides the number

⁴² Linguistic typology fails to constitute a homogenous field, since no coherent results are produced. Therefore, it ought to be viewed as a sequence of different approaches rather than an ultimate language explanation (Luschützky 1999).

of categories which equals the number of phenomena to be categorized. However, a typology providing as many types as there are phenomena to categorize misses relevant generalizations and is not optimal at all. Dualistic typology seems to be particularly problematic since it draws a distinction between abstract *good* and *bad* types (Luschützky 1999). The seemingly dualistic typology of processes which operates with *the speaker-friendly* and *hearer-friendly* categories involves subtypes since both types are *good* in certain respects and at the same time *bad* in others. Furthermore, a dualistic, *black-and-white* typology may produce either positive or negative results, leading to both insights and mistakes. Suffice it to mention the cases of infelicitous typologies which, for instance, classified the representatives of white race as superior to the representatives of the black race. This case demonstrates that a *black-and-white* typology may give rise to prejudices and create stereotypes. Therefore, any proposed typology should assume and promote a detached attitude towards entities it categorizes. Otherwise, the categories do injustice to the entities they describe. Thinking based on categories is pervasive: once a category is ascribed to an entity, it takes a long time to undo the process. Moreover, the semantics of the categorized types (e.g. negative connotations of the adjective *black*) interferes with a metalanguage of typology.

Linguistic typology must pay regard to the difference between the external and the internal properties or features. People consider languages to be official or minority languages, which is based on external factors (political implications of the language-dialect distinction). Another common distinction between *melodic* versus *monotonous* (e.g. tone languages versus prosodically *flat* languages) serves as an example of typology based on internal properties (Luschützky 2005). However, the above distinction is not scientific since it mixes up superficial, impressionistic properties with the structural ones. Genuine linguistic typology employs exclusively language internal properties which are grammatical. In fact, all components of the linguistic system must be taken into consideration (e.g. phonetics, phonology, morphology, syntax, semantics etc.).

6.3. The current approaches to lenition/fortition

The typology of phonological processes has an extensive literature. The two types of phonological processes, i.e. fortition and lenition, have incited much interest within the linguistic science since its inception. The history of studies in phonological processes can be traced back to the Sanskrit grammarians several centuries B.C. (Sharma 1987, 1990). The 19th century witnessed a genuine peak of interest in phonological processes. This interest stemmed primarily from the advent of diachronic studies. The first formalized approaches to process typology were these of Grimm (1918), Bopp (1863) and Curtius (1856-1862). These linguists have pioneered the field of process typology and commenced a systematic analysis of phonological processes in terms of types. In their works, the following process types have been identified: assimilation, dissimilation, absorption, epenthesis, metathesis, haplology, syncope and apocope (Luschützky 1997). It was Sievers (1876) who introduced the distinction between sound substitutions (*Lautwechsel*) and sound changes (*Lautwandel*). A number of issues related to process typology have been taken up and discussed at length in the 20th century. The modern studies owe much to the advent of generative grammar (Ferguson 1978). The generative approach, however, abandoned the process studies in the course of development in favor of rules. Nowadays, the nature and typology of phonological processes are the major interests of Natural Phonology (Donegan – Stampe 1979, Dressler 1985a, 1985b), Modern Natural Phonology (Dziubalska-Kołodziejczyk 2003, 2004) and Optimality Theory (Boersma 1998, Kirchner 1998, Jun 2004). The latter approach explores mostly lenitions in terms of constraints, not processes. Therefore, the abovementioned approaches are revised in the following subsections due to their direct link with phonological processes.

6.3.1. The traditional approach

The typology of phonological processes is dualistic. There exist two types of phonological processes: lenitions and fortitions which are differentiated on the basis of the strength of a sound, or energy expended in its production. Voiced sounds are called lenis (weak), whereas voiceless ones are called fortis (strong). Malmberg (1963) explains and exemplifies this distinction in terms of airstream resistance:

A consonant may be articulated with more force or less force. The current of air can be more or less intense. And the resistance offered to the current of air at the point of articulation of the consonant may be more or less vigorous. There are *strong* consonants and *soft* consonants. In English or French, the voiceless stops [p], [t], [k], and the voiceless spirants [f], [s], and [ʃ] are *fortes*, their voiced counterparts [b], [d], [g], [v], [z], [ʒ] are *lenes*. Among the stops and spirants, there are thus two series – a series of fortis which are opposed to a series of lenes ([p] to [b], [s] to [z], etc.) The nasals and liquids, as also the “semivowels”, in European languages are always lenes (Malmberg 1963: 51-52).

The fortis/lenis distinction derives from the greater/lesser pressure of air built up under the vocal folds which, in turn, results in the greater/lesser force of articulation. Bussmann (1996) provides the following definition of the greater/lesser force of articulation:

[A]rticulatory feature of **stops** and **fricatives** that refers to differing degrees of muscle tension. In fortis sounds, the subglottal air pressure behind the point of articulation is stronger than in lenis sounds. The partially synonymous terms **tenuis vs media** refer only to stops and denote that aspect of voicelessness vs voicedness [...] that correlates within the features [fortis] vs [lenis] in English. Moreover, the fortis/tenuis sounds [p, t, k] in English are aspirated [...] to varying degrees depending on their position in the given word (e.g. word-initial, word-medial, word-final) (Bussmann 1966: 171).

The dualistic typology of processes reflects the force of articulation and involves its modification. Therefore, the processes of the lenition type substitute the fortis sounds with the lenis ones, whereas the processes of the fortition type substitute the lenis sounds with their fortis counterparts. Trask (1996) defined fortition in the following way: “(also **strengthening**). Any phonological process in which some segment becomes ‘stronger’ (more consonant-like). An example is the development of the glide [j] into some kind of fricative, affricate or plosive in most varieties of Basque” (Trask 1996: 149). The above definition highlights the nature of the fortition processes which affect the lenis sounds, transforming them into the fortis ones. The factors triggering lenition have been widely identified as a change towards ease of production and a reduction in the degree of articulatory complexity via weakening. Bussmann (1996) discusses the properties of the lenition processes and stresses the distinction between the consonantal and vocalic lenition:

Phonetically motivated process of sound change that leads to the reduction of sounds and, in extreme cases, to loss of segments; typically this occurs in positions where assimilation is favored or in syllabically ‘weak’ positions (e.g. in final position, in unstressed syllables). Two types of weakening are distinguished. (a) Consonant weakening (also lenisization):

this denotes a weakening of **consonant** strength (through a reduction in air pressure and muscle tension or an increase in sonority) to the complete loss of a segment [...] (b) vowel weakening: this is a term for all processes that lead to a weakening of the articulatory movement in the sense of an increasing centralization of vowels and finally a total loss of the vowel (Bussmann 1996: 519).

Trask (1996) defines lenition as follows:

([A]lso weakening). Any phonological process in which a segment becomes either less strongly occluded or more sonorous, such as [k] → [x], [x] → [h] or [k] → [g]. Often the term is extended to various other processes, such as loss of aspiration, shortening of long segments and monophthongization of diphthongs, which represent ‘weakening’ in some intuitive sense (Trask 1996: 201).

The above definition implies that the lenition/fortition are defined in a vague, intuitive way and fail to receive a full-fledged specification. This implication extends to the whole traditional approach to process typology which incorporates the strength of a sound and the force of articulation as the lenition/fortition characteristics.

6.3.2. The NP approach

In the NP literature, phonological processes are divided into lenitions and fortitions on the basis of the functions they serve and the context in which they appear. Both lenition and fortition operate on a segmental level, as opposed to prosodic processes which are located at a suprasegmental level (Luschützky 2001). Moreover, their labels refer to various aspects of language: centrifugal/centripetal refer to the phonetic space, strengthening or weakening refer to phonetic gestures whereas foregrounding or backgrounding address communicative teleology (Luschützky 2001).

Fortition processes, also referred to as strengthening or centrifugal, perform the listener–friendly function. Since fortitions strengthen the clarity of perception, they enhance contrast for the sake of a better, sharper perception. They have a perceptual teleology. They operate independently of the context (rely on the system inventory) and are style-sensitive (appear in formal/lento/emphatic speech). The operation of fortitions consists in affecting the segments in strong positions. The nature of fortitions is paradigmatic due to the fact that this type operates on individual sound segments (Donegan – Stampe 1979). Lenitions, also known as the weakening or centripetal processes, perform the speaker–friendly function. Their teleology is rooted in

articulation as lenitions serve the ease of articulation, lessening the pronounceability burden on the part of the speaker. Lenitions are sensitive to context which follows from their syntagmatic nature. They operate on sound sequences which must observe the context. Lenitions arise in the speech styles which invite ease of articulation: informal/allegro/fast/rapid (Donegan – Stampe 1979). The lenition/fortition processes are also expressed in terms of semiotics. Dressler (1985a) highlights the clarification function of fortitions and terms them the *foregrounding processes* (Verdeutlichung: of dissimilatory nature). Lenitions which serve the obscuration function are termed the *backgrounding processes* (Entdeutlichung: of assimilatory nature). The terms *foregrounding* and *backgrounding* derive from the semiotic figure-and-ground principle which “sharpens the contours of what is to be perceived” (Dressler 1985a: 3). In turning to semiotics not only does NP derive the lenition/fortition criteria from the functions they serve, but also from the process place in lexicon. Another way of describing processes, prelexical and postlexical, situates processes with reference to lexicon (Dressler 1985a). The prelexical processes define the phoneme inventory (segment structure) and govern the phonotactics of phonemes (sequential structure). The postlexical processes “derive phonetic output from phonemes” (Dressler 1985a: 30). Prelexical processes are fortitive, whereas postlexical are lenitive. Dressler (1985a) noted that “the same process may function both pre- and postlexically” (Dressler 1985: 30). Thus, he predicted that lenition/fortition and the post/prelexical characters “will coincide largely within the phonology of any language” (Dressler 1985: 33). In terms of lenition/fortition typology, Luschützky (2001) proposes categories as the basis of process typology such as phonetic (strengthening/weakening), structural (paradigmatic/syntagmatic), grammatical (prelexical/postlexical) and functional (adaptive/evolutive) (Luschützky 2001).

Within NP, phonological processes with reference to consonants were studied in depth (Goman 1979):

Fortitions create phonology. They not only refer to our perception of the speech act, they also account for it. Fortitions regulate what sort of thing can count as a mental representation, or mental intention, concerning speech. Lenitions, on the other hand, lead to phonetics. They regulate our notion about what is a suitable or affordable utterance (Goman 1979: 43).

6.3.3. The OT approach

The OT formalism does not incorporate lenition/fortition as phonological processes, but rather as universal and violable constraints which compare candidates or candidate forms. Nevertheless, what is considered to be the typology of phonological processes, receives considerable amount of interest within OT. Regarding the discussion of phonological processes, the most influential approaches in this theory are these of Boersma (1998) and Kirchner (1998, 2004). In particular, the issue of lenition was taken up by the two authors. Under their approaches, lenition denotes the group of phonological processes which is driven by a natural need to minimize articulatory effort and therefore, is effort-based. Effort is formalized as a constraint (the LAZY constraint in Kirchner 1998 and the MIN EFFORT constraint in Boersma 1998). Articulatory effort is employed in the sense of biomechanical parameters such as precision, distance, coordination energy, mass etc:

In order to develop a concrete effort-based analysis of the foregoing lenition patterns, it is necessary to make explicit assumptions concerning the relative effort costs of the relevant consonants [...] in terms of a set of inferences [...] concerning relative effort cost of particular consonant types, which follow largely from the equation of effort with biomechanical energy (Kirchner 2004: 320-321).

It appears from the above quotation that the OT effort-based approaches attach the utmost weight to effort in the biomechanical sense. Jun (2004) offered an effort-based treatment of place assimilation. She formalized the WEAKENING constraint in terms of minimization of articulatory effort which she derives from articulatory and perceptual domains. Jun (2004) noted: “WEAKENING thus has the effect of reducing or eliminating articulatory gestures, leading to place assimilation in consonant clusters” (Jun 2004: 59). On the basis of the constraint she developed a typology of place assimilation, including target manner, target place, position of target and trigger place.

Another approach to lenition within OT in terms of autosegmental feature spreading was proposed by Selkirk (1980) and Harris (1984). The former author argued that in Spanish the lenition process of spirantization consists in spreading the feature [+continuant]. This approach, however, fails to capture the fact that deletion of any phonological material (i.e. the nature of lenition) bars any features from spreading (Kirchner 2004). Another approach to lenition was that of Foley (1977) and Clements

(1990) who pursued the notion of lenition as scalar sonority promotion (along the sonority scale). The sonority scale on which the sounds are organized according to the energy relative to effort has the following order:

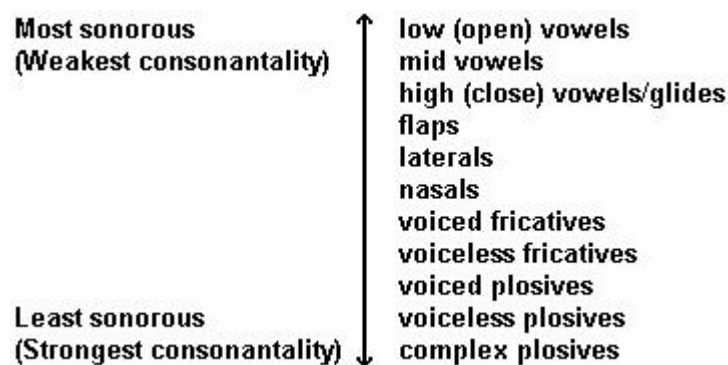


Figure 39. An illustration of sonority scale (after SIL International 1999).

Sonority promotion is implicational. If plosives lenite to fricatives, then the order predicts that spirants should lenite to nasals. This prediction, however, fails to find support in the phonological facts (Kirchner 2004), whereas spirantization has been attested (Blevins 2004). Foley himself (1977) provides two definitions of lenition/fortition: “the phonological weakening of elements as opposed to *strengthening*, the phonological strengthening of elements” (Foley 1977: 143) and “lenition in its conversion of elements to weaker elements also indicates relative phonological strength” (Foley 1977: 44).⁴³

6.3.4. The evaluation of the current approaches to process typology

Regardless of the theoretical approach, the very terms *lenis* and *fortis* as well as *weakening* and *strengthening* appear problematic since they raise doubts among various linguists. The doubts have been voiced with various frequency and intensity. For instance, Lisker (1972) expresses the following view: “If [...] it is supposed that the two stop categories differ in force of articulation only where they differ in duration, then the *fortis-lenis* difference no longer can be said to fulfill the distinctive feature role as this is

⁴³ It must be specified that this work of Foley (1977) is affiliated with transformational grammar rather than OT as he states: “Theoretical phonology allows a formalization of the traditional concept of lenition as $\alpha_1\beta_2 \rightarrow \alpha_1\beta_1$ which captures the characterization of lenition as a weakening (reducing the β strength from 2 to 1), while the transformational phonetic formulation [...] makes no reference to weakening, thus failing properly to characterize the process” (Foley 1977: 29).

usually understood” (Lisker 1972: 342). In similar vein, Luschützky (2001) criticizes weakening/strengthening as one-dimension, abstract phonological parameters:

It is clear that our strength scales are abstract in the sense that they are restricted to categorical gestures, i.e. phonotactic actions generating phonological classes of sounds. In addition to that, there is of course a purely phonetic kind of general strength. For instance, pulmonic initiation can be gradually weakened or strengthened according to the overall loudness of utterances (Luschützky 2001: 508).

Rather, Luschützky (2001) considers sounds in terms of gestures whose states must be analyzed separately. Yet another example of distrust in weakening/strengthening in its conventional terms is provided by Blevins (2004):

The role of perception in leniting sound change creates problems for restrictive phonological account. Lenition cannot be modeled as phonological feature loss or delinking (Harris 1997), with subsequent spreading of marked features (Kiparski 1988), since intervocalic voicing and intervocalic spirantization involve insertion of specified feature values [+voiced] and [+continuant], which, within constrained theories of underspecification, should be unspecified for vowels. Nor can phonological lenition always be characterized as a decrease in effort (Kirchner 2000), since voicing a stop arguably requires more effort than not. See Lavoie (2001, chapter 6) for similar assessments (Blevins 2004: 147).

In similar vein, Pierrehumbert – Talkin (1992) having commented the studies of /h/ and /ʔ/ in different positions in a word or a phrase in Korean arrive at the following conclusion regarding lenition: “A quantitative articulatory description will still fail to capture the multidimensional character of lenition if it handles only local phonological and phonetic properties” (Pierrehumbert – Talkin 1992: 117). Similarly, Watson (2006) studies lenition in Liverpool English and identifies the following problems with lenition:

The lack of agreement amongst linguists on what lenition is perhaps surprising given the pervasiveness of the discussion, but it is nevertheless apparent [...] Some of the first explanations usually offered for what lenition is are centered around phonological strength. The definition of phonological strength (i.e. that certain segments can be inherently stronger or weaker than others) is problematic because strength is rarely tied to any acoustic or articulatory correlate (e.g. Foley 1977) and instead is related to the relationship and interaction between segments (Watson 2006: 32-33).

Lenition broadly understood operates in the direction from a more to less difficult sounds. Under the traditional, most common approach (e.g. Trask 1996, Bussmann 1996), lenition substitutes the fortis sounds with the lenis ones, whereas fortition substitutes the lenis sounds with their fortis counterparts. Such a definition is circular instead of explanatory. This lenition/fortition definition is based on the strength of the sound and reflects only the force of articulation. It operates in an automatic, indiscriminate way, which does not take into account the mental character of phonological processes. It also stipulates that the speaker is not aware of the processes implemented in his or her own speech. Moreover, fortition is just the reverse operation along the fortis/lenis scale. This type of fortition definition raises certain doubts since in terms of frequency, lenition processes constitute the majority, therefore their role and origins should at least be different from fortition but not deriving from its reverse.

Within the NP framework, the lenition/fortition definition is based on the needs of the speaker as well as the listener and offers an operational procedure: if the phonological material is deleted for the benefit of the speaker, it is a lenition; when the material is added for the sake of the listener, it is a fortition. Furthermore, according to the common belief held widely in linguistics and in NP the roles of the speaker and the listener are of equal importance. They are not since in the communication act it is the speaker who must satisfy both his or her own needs as well as the needs of the listener. The significance of the roles performed by the speaker and the listener is not really equally distributed since the speaker, unlike the listener, deals with pronounceability and perceptibility at the same time, bearing the entire burden of articulation.

The OT approach (Boersma 1998, Kirchner 1998, 2004) advocates articulatory effort as the motivation of lenition and fortition. For instance, lenition is effort-based and driven by a natural need to minimize articulatory effort (Kirchner 1998). Articulatory effort is employed by Boersma in the sense of biomechanical parameters such as precision, distance, coordination, energy, mass etc. There is no denying that these parameters can be measured. Moreover, a holistic approach could be implemented, under which the parameters can be simply added (in fact, this approach was adapted in the first experiment, cf. 4.4.1.). It would also be interesting to establish the role of individual parameters in the overall effort. The role could be resolved by manipulating (adding, deleting, increasing and decreasing) the parameters and one could judge that precision, out of remaining several parameters, accounts for say 10 per cent, not 90 of the overall effort. The holistic approach, however, was merely signaled

in the OT literature (Boersma 1998) but failed to become a standardized measure. Thus, the idea of biomechanical parameters as such is not subject to critique, unlike the lack of the idea's implementation. Besides, biomechanical parameters as the solely lenition criteria do not take into consideration the mental reality of processes.

The evaluation of the current approach to lenition/fortition leads to the following observations. First, there is no exhaustive definition of lenition/fortition whereas the existing ones are either circular in the traditional approach (e.g. Trask 1996, Bussmann 1996) or automatic in the current approaches to lenition/fortition, i.e. they are based on the erroneous assumption that phonological processes are automatic, a mere substitution of weak sounds for the strong ones or an indiscriminate deletion/addition of sounds. If it were true, all languages of the world would be the same and this is simply not the case. Second, in the absence of a satisfactory definition of lenition or fortition it is still not clear what classifies a given process as lenition or fortition. Third, lenition is extensively covered but there are not equally numerous studies on fortition. As a matter of fact, it seems that only Goman (1979) directly addressed the issue of fortition in consonants, whereas typically, fortition is mentioned as the reverse of lenition and not studied in its own rights. Fourth, the current approaches classify processes as lenition/fortition on the basis how a process operates, not on what it does (procedure is considered, but its result is not taken into account at all). Evaluating the current debate on phonological processes, it appears that although lenition and fortition have an extensive literature, a number of controversial issues can still be identified. Having revised the current approaches to lenition/fortition, it can be concluded that none of them manages to present a satisfactory definition of lenition/fortition, whereas their typology remains problematic.

The above discussion reflects the degree of confusion in the areas of phonological processes and process typology. Not only do current approaches define lenition/fortition or specify their typology in a vague, circular way but also fail to acknowledge the mental reality of processes which is postulated in NP (Donegan – Stampe 1979). Dziubalska-Kołaczyk (2006) expressed the concern about the lack of explicit specification of lenition/fortition in MNP.

6.4. The definition and criteria of lenition and fortition proposed in the thesis

Having identified the two issues in the area of phonological processes (the issue of defining lenition/fortition) and process typology (the issue what classifies a given process as lenition or fortition), the following subsections address them.

6.4.1. Definition of lenition and fortition proposed in the thesis

In order to address the definition issue, the very idea behind lenition/fortition must be reconsidered. First, the definition of lenition is revisited. Lenition is a speaker-friendly process and as such consists in reduction of “a specific common difficulty to the speech capacity of the individual” (Stampe 1973 [1979]:1). The question remains what exactly is reduced. Undoubtedly, lenition reduces energy in terms of lenis/fortis categories. These categories originate from phonetics. Energy, however, provides a static understanding of lenition, offering a fortis/lenis scale on which the speaker moves to left or right from the point zero. Besides, energy of a sound works well when a sound is considered in isolation. When context is applied, energy becomes relative. Moreover, energy itself does not suffice since it is not the only factor of speech that is subject to modification. Energy as a static and relative criterion of lenition is just one dimension of lenition, whereas the dynamic aspect of speech is neglected. Therefore, the definition of lenition should also incorporate gestures (cf. Chapter Three). Luschützky (1997) incorporated the notion of gestures in his typology of phonological processes, therefore this line of thinking should be continued and applied to the definition of phonological processes. Moreover, the aerodynamics of the vocal tract ought to be taken into account. Incorporating energy, gestures and aerodynamics of the vocal tract results in three criteria of lenition and from the three criteria, a negative *definition* of fortition can be inferred (Katarzyna Dziubalska-Kołaczyk, p.c.). The three criteria based on the understanding lenition as reduction were expressed in *negative* terms/labels (energy, gestures, aerodynamic unnaturalness) since they are subject to reduction:

- a) Energy: lenition viewed as the speaker-friendly process operates in the direction from more to less energy, i.e. from fortis to lenis sounds. For instance, voicing of voiceless obstruents involves the loss of energy: (p → b). Thus, lenition can

be understood as reduction of energy when the fortis/lenis categories are applied.

- b) Gestures: lenition as the process which reduces the speaker's task operates in the direction from more to less complex sounds. Complexity consists in the number of articulatory gestures or features required for a given sound. Therefore, complexity is calculated paradigmatically, i.e. the context is not taken into account since even in sound sequences the neighboring sounds do not change the gestural make-up of a sound. Reduction means fewer articulatory gestures or fewer features. The following processes exemplify lenition as reduction of complexity: vowel centralization ($\text{æ} \rightarrow \text{ə}$), monophthongization ($\text{ʊə} \rightarrow \text{ɔ}$), assimilation of place ($-\text{np}- \rightarrow -\text{mp}-$), devoicing of final obstruents ($\text{b} \rightarrow \text{p}$) and deletion (*last night* $\text{læst naɪt} \rightarrow \text{læs naɪt}$). Thus, lenition can be understood as reduction of complexity when the notions of gestures and gestural complexity are applied.

Under traditional definition of lenition/fortition, voicing ($\text{p} \rightarrow \text{b}$) is viewed as lenition, whereas devoicing ($\text{b} \rightarrow \text{p}$) as fortition. It must be clarified why these seemingly contradictory processes such as voicing and devoicing are subsumed under lenition. If lenition is understood as complexity reduction, then it manifests itself as the process of devoicing ($\text{p} \rightarrow \text{b}$). If lenition is understood as energy reduction, then it manifests itself as voicing ($\text{b} \rightarrow \text{p}$). These two understandings of lenition as complexity or energy reduction should not be perceived as mutually exclusive or contradictory. More complexity does not necessarily mean more energy: /b/ is more complex than /p/ due to the fact that articulation of /b/ involves one articulatory gesture more (voicing). On the other hand, /b/ requires less energy than /p/ due to the more vigorous resistance offered by the narrowed configuration of the vocal folds to the current of air. Therefore, the articulatory complexity of a sound is independent of the energy expended in the articulation, and the other way around.

- c) Aerodynamic unnaturalness: lenition operates in the direction from less natural sounds to more natural sounds. Aerodynamic unnaturalness, unlike gestural complexity, is calculated syntagmatically, i.e. the context is taken into account. In this case, lenition is driven by the structure of the vocal tract and the aerodynamic processes which apply to the movements of the articulators. This

particular understanding of lenition is the most rudimentary, low-level phonetic as it derives from the design of the vocal tract. The following processes exemplify lenition as reduction of aerodynamic unnaturalness: the *prince* epenthesis (*prince* prɪns → prɪnts) and final devoicing (b → p).⁴⁴ Thus, lenition can be understood as reduction of aerodynamic unnaturalness when aerodynamics and motorics of the vocal tract is applied.

Summing up, the definition of lenition proposed in the thesis is as follows: lenition is reduction of the three criteria: energy, complexity or aerodynamic unnaturalness. It must be emphasized that the criteria are not of equal significance. For instance, in assimilation processes the criterion (c) itself is not sufficient, aerodynamic unnaturalness is reduced first, then the criterion (b) is applied as assimilations also happen to reduce the number of gestures (e.g. in *good book* aerodynamic unnaturalness is reduced via place assimilation since place of articulation varies, then one of the two /b/s is reduced via gestural reduction). It seems that aerodynamic unnaturalness reduction necessitates lesser gestural complexity. Therefore, the three criteria of lenition can be interpreted in various ways in terms of their order:

- a) Reduction of energy: the interpretation of this lenition criterion is not unique, this type is relative as this lenition criterion depends on context.⁴⁵ In general, according to energy criterion /b/ is less effortful than /p/ in isolation, but in one context /b/ is better (in *web domain* /b/ should not be devoiced as it precedes another voiced sound), whereas in another, e.g. *web traffic* /p/ is better (cf. Chapter Three, section 3.2.). Moreover, fortis/lenis categories are operational within the voiced/voiceless categories but not outside them which excludes this lenition type in processes involving two sounds from the two categories such as assimilations (e.g. where voiceless /t/ is assimilated to voiced /j/ in *don't you*) except for assimilation of voicing.
- b) Reduction of gestures: the interpretation of this lenition criterion is unique, it is absolute and autonomous. It must be considered in terms of segments

⁴⁴ The *prince* epenthesis is the type of a phonological process where a voiceless stop/fricative is inserted before a nasal or lateral which neighbors /s/, as opposed to, e.g. vowel insertion which is also known as epenthesis.

⁴⁵ Cf. Chapter Five, section 5.6. in which Luschützky (1997) is reported to call the fortis/lenis categories relative opposite positions on the grounds that they fail to account for absolute positions in perception.

(paradigmatically), on the level of gestures of sounds only and not on the sounds sequences. Regardless of the context, the gestural make-up of a sound never changes. /p/ is less complex gesturewise as it always has one gesture less than /b/. Lenition according to Scheer (Tobias Scheer, p.c.) is determined by its position in a word, but if one considers what is reduced (energy or gestures) it seems that the absolute lenition criterion is not determined by its position in a word because its segmental make-up does not depend on position.

- c) Reduction of aerodynamic unnaturalness: the interpretation of this lenition criterion is not unique, this criterion is relative. It depends on context which creates aerodynamic unnaturalness. It must be considered in terms of sound sequences (syntagmatically) but it must also take into account the result and not only the operation of a process since fortitive processes have lenitive effects (e.g. epenthesis).

In terms of criteria description, lenition criterion (a) (energy reduction) and lenition criterion (c) (aerodynamic unnaturalness reduction) are relative and context-dependent, whereas lenition criterion (b) (complexity reduction) is absolute. Criteria (b) and (c) operate on two planes: syntagmatic and paradigmatic. In terms of the ordering of lenition criteria, the following hierarchy can be proposed:

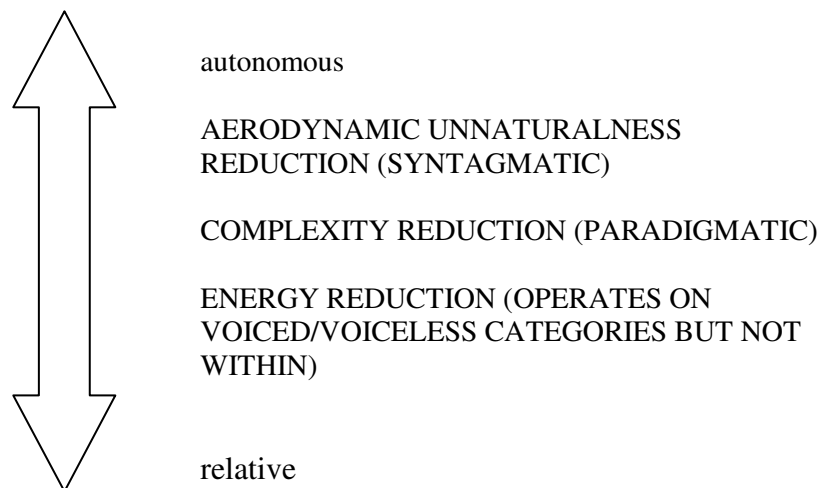


Figure 40. A hierarchy of lenition criteria.

Figure 40 arranges the three lenition criteria on a scale. Aerodynamic unnaturalness reduction is at the top of the scale, whereas complexity reduction follows it. It must be stressed that although both criteria complement each other like it is the case of assimilation, criterion (c) has primacy over criterion (b) on the grounds that the

structure of the vocal tract is absolute, whereas complexity comes from a system. Regardless of a system inventory, the vocal tract does not vary across languages. Energy reduction is at the bottom of the scale as it is the relative criterion and fails to prove effective outside voiced/voiceless categories. It must be emphasized that although both criteria (b) and (a) concern the ways of articulating a sound (with less/more energy or greater/lesser gestural complexity), criterion (b) has primacy over criterion (a) on the grounds that the number of gestures in a given sound cannot be manipulated, unlike energy. Thus, the hierarchy is organized by autonomousness of the criteria and the next step, which unfortunately cannot be taken in the absence of gestural sound descriptions, would be to establish the relations between them. Aerodynamic unnaturalness depends on context but not on the system, gestural complexity depends on the system but not on the context, energy depends on the context and operates only within voiced/voiceless categories. It must be emphasized here that the hierarchy is not absolutely strict since the criteria can be also viewed as are preferential constraints.

These three hierarchically organized criteria of lenition (lenition understood as their reduction) give rise to a negative definition of fortition. The following example of a phonological process shall explain why the definition of fortition is negative: epenthesis represents the understanding of lenition as reduction of aerodynamic unnaturalness. This process is described as a transient desynchronization of the nasal and oral, say alveolar, gestures of closure. Desynchronization itself, however, cannot serve as an explanation. It is purposeful and meaningful. Although the description allows the fortition typology, the motivation of the desynchronization clearly points to its lenitive nature. Desynchronization means the reversal of the usual order in which denasalization precedes the release and a short interval of oral closure results (i.e. the consonant such as /t/ is produced). Therefore, it is logical and natural to insert that stop. An alternative understanding of fortition follows from this particular example. Since the pronunciation of [prints] reflects the structure of the vocal tract, the pronunciation of [prɪns] would require an effort to suppress a natural outcome of motorics and aerodynamics. In fact, more articulatory effort (epenthesis) must be expended in order to obtain the optimal result: not to have to repeat and to be understood immediately. This implies that fortition is a suppression of lenition. The suppression resembles the effort expended in the *unlearning* as the English children must unlearn to naturally devoice the final obstruents. Thus, the order of first language acquisition provides

support to the negative definition of fortition. As a matter of course, one must draw the distinction between process and process state. Under the thesis' approach, fortition negatively defined does not represent a process state since its operation is aimed towards suppressing. If one defines process as action and process state as the lack of action, fortition being suppression of lenition consists in an action.

As far as lenition motivation is concerned, the speaker is the source of lenition. Lenition is a tendency to decrease effort, whereas fortition lacks this motivation completely. Fortition manifests itself in e.g. repetitions and citation forms and effort is lenition suppressing-oriented. If one assumes that casual speech is normal speech (Shockey 2003), then lenition is a default option for the speaker. The speaker may get closer or nearer that default option. Therefore, if the speaker uses fortition, effort is expended in getting away from the default option of lenition, i.e. increases in order not to apply lenition.

Summing up, the thesis proposes to define lenition as reduction and proposes its three criteria: reduction of energy, reduction of complexity and reduction of aerodynamic unnaturalness (they can be ordered on a scale, cf. Figure 40), whereas fortition is an effortful suppression of lenition. In terms of motivation, lenition is the speaker's default option. The advantage of these definitions is that they acknowledge the dynamic aspect of speech in terms of gestures and motorics of the vocal tract, proposing other criteria of lenition than energy. The definitions also provide processes with mental reality since the speaker mentally deactivates the lenition option in the case of fortition.

6.5. Revision of process typology under lenition criteria

Regardless of the fact that the definitions of lenition/fortition lack an explicit characterization, examples of individual processes are discussed at length. Unfortunately, no exhaustive compilation of processes exists in the linguistic literature, presumably due to the fact that each theory investigates only selected aspects of processes and selected examples are provided. Nevertheless, the phonetic conditions for a particular process seem to be extensively covered (Dziubalska-Kořaczyk 2003,

Dressler 1995, Kirchner 1998, Blevins 2001, Gurevich 2004), whereas there is a marked tendency that the complete list of the processes is merely signalized. Table 10 presents the processes discussed by various authors. Table 10 reflects the current approaches (the name, relevant source and, if possible to establish, theoretical affiliation indicated). Each and every particular phonological process is discussed below in order to justify the revised typology (Tables 11 and 12) and in order to consider the understanding of lenition in question.⁴⁶

Table 10. The current typology of phonological processes.

Fortition	Lenition
Diphthongization (Dressler 1995: NP)	Monophthongization
Epenthesis (Dressler 1995: NP)	Shortening (Dressler 1995: NP, Kirchner 1998: OT, Mateescu 2003)
Vowel insertion (Dziubalska-Kořacyk 2003: MNP)	Weakening: fricativization, gliding (Dressler 1995: NP, Kirchner 1998: OT ⁴⁷)
Lengthening (Dressler 1995: NP)	Centralization (Dressler 1995: NP)
Strengthening: stopping, aspiration (Dressler 1995: NP)	Segment deletion (Dressler 1995: NP, Dziubalska-Kořacyk 2003: MNP, Kirchner 1998: OT)
Devoicing (Kirchner 1998: OT, Mateescu 2003)	Cluster reduction/simplification (Dziubalska-Kořacyk 2003: MNP)
Dissimilation	Assimilation of stops and nasals (Dressler 1995: NP, Dziubalska-Kořacyk 2003: MNP, Jun 2004 :OT)
	Palatalization: Yod coalescence (Dziubalska-Kořacyk 2003: MNP)
	Degemination (Dziubalska-Kořacyk 2003: MNP, Kirchner 1998: OT)
	Hiatus avoidance via linking or intrusive /r/ (Dziubalska-Kořacyk 2003: MNP)
	Smoothing (Dziubalska-Kořacyk 2003: MNP)
	Flapping (Kirchner 1998: OT)
	Debuccalisation (Kirchner 1998: OT)
	Voicing (Kirchner 1998: OT)

⁴⁶ Since the three independent understandings of lenitions are proposed, it is necessary to illustrate these types at work.

⁴⁷ Gurevich (2004) in her cross linguistic study investigates a set of 153 languages. In fact, the set comes from Kirchner (1998), who, in turn, based his database on the one compiled by Lavoie (1996, 2001). In similar vein, Blevins (2004) also uses the database of Lavoie. Therefore, it was considered appropriate to indicate Kirchner (1998) in the table as the author of the list instead of Gurevich since the processes were taken from his work.

As a matter of course, the processes should be analyzed from the angle of presence, absence and correlation between the three lenition criteria. Given the fact that there is no phonological literature which would describe sounds and sequences of sounds in English in terms of articulatory gestures or the transitions between the gestures, it is not possible to establish how a given lenition criterion behaves in a given process. In the absence of the descriptions, the following revision must be limited to the criteria or criterion which is the most evident and comes forward. It was possible, though, to establish how effort is distributed in each of the process, thus the revision indicates lenition criterion as well as the distribution of articulatory effort.⁴⁸ All processes are revised in a teleological way, i.e. according to their result, not operation (cf. section 6.7.) since the definitions of lenition as segment deleting or fortition as segment adding do not take into account mental reality of processes or their multifunctionality (many forms can serve the same function), stipulated in NP. In order to facilitate revision, the processes were grouped according to their umbrella terms (i.e. I. reduction/deletion II. Insertion III. duration change IV. assimilation V. substitution).

I. PROCESS GROUP: REDUCTION/DELETION

6.5.1. Segment deletion/ cluster reduction/simplification

This process stands for reduction of any consonantal cluster by one segment (e.g. *stand* by stænd baɪ → stən baɪ). Three consonants in a sequence require greater complexity than two consonants in a row. Since the workload for articulators is reduced, the results of this process qualify as lenition since gestural complexity is reduced, effort is thus reduced.

6.5.2. Degemination

This process type stands for reduction of a long (geminate) to a short (singleton) consonant (e.g. *bad drive* bæd draɪv → bædraɪv). One gesture is required instead of

⁴⁸ Effort is not referred here to as articulatory effort exclusively, but also as supression of lenition.

two and this kind of reduction meets the conditions for lenition as complexity reduction, effort being reduced.

6.5.3. Monophthongization

This process type stands for reduction of a diphthong to a monophthong (e.g. *time* taɪm → ta:m). This means one articulatory gesture instead of two. The process can be typologized as lenition since gestural complexity is reduced, effort is thus reduced.

6.5.4. Smoothing

This process stands for reduction of the vowels number in a sequence such as triphthongs (e.g. *layer* leɪər → le:ər). The process can be typologized as lenition since gestural complexity is reduced, effort is thus reduced.

6.5.5. Centralization

This process stands for reduction of any vowel to schwa (e.g. *and* ænd → ənd). The central position assumed by the tongue is the most “natural” and effortless since the tongue does not have to reach any targets required by height or position of the tongue. The higher the degree of vowel centralization, the less effort invested. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided. It also can be typologized as reduction of gestural complexity since schwa makes all the gestures such as aiming for the proper height and assuming the correct position of the tongue for a given vowel redundant.

II. PROCESS GROUP: INSERTION

6.5.6. Epenthesis

This process stands for an insertion of a vowel or a consonant (cf. section 6.4.). In English, there is a process of consonantal epenthesis which inserts homorganic voiceless stop after nasal and before voiceless fricative followed by an unstressed vowel in the same word (e.g. *prince* prɪns → prɪnts). The presence of /t/ increases the number of articulatory gestures. The insertion of a consonant is viewed as a lenition understood as an outcome of aerodynamics and motorics of the vocal tract. In the course of the transition from /n/ to /s/, the three following features change: the state of the glottis (from voiced to voiceless), manner of articulation (from stop to fricative) and state of the velum (from nasal to oral). Thus, the production of /t/ results as a side-effect and serves as the bridge which helps to overcome the difficulties of transition. The analysis of consonantal insertion as a lenition derives from aerodynamics of the articulators. Lenition stirs numerous controversies. These controversies arise from the fact that the description of the process is mistaken for its explanation. Epenthesis is described as a transient desynchronization of the nasal and oral, say alveolar, gestures of closure. Desynchronization itself, however, cannot serve as an explanation. It is purposeful and meaningful. Although the description points to a fortitive process, the motivation of the desynchronization clearly points to its lenitive nature. Desynchronization means the reversal of the usual order in which denasalization precedes the release and a short interval of oral closure results (i.e. the consonant such as /t/ is produced). Therefore, it is logical and natural to insert that stop. Failure to do so, or the suppression of insertion, is possible but is more effort-costly. The recent study by Oñederra – Jauregi (2005) demonstrated that epenthesis is governed by the syllabic and rhythm pattern of a language. In Basque, desynchronization which results in epenthesis triggers affrication. Sibilant fricatives in the syllable onset preceded by a sonorant consonant are affricated: *abertzale* ‘patriot’ (*aberri* ‘country’ + *zale* ‘lover’) vs. *etxezale* ‘homelover’ (different context). The affrication arises in Basque due to the fact that the syllable structure creates more favorable conditions than in English. The rhythm of Basque facilitates the process which optimizes the syllable structure in terms of head strengthening (Oñederra – Jauregi 2005). The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus expended.

6.5.7. Vowel insertion

This process stands for the violation of phonotactic constraints of English by non-native speakers. For instance, the Polish speakers of English insert vowels in the clusters they consider ill-formed and difficult to pronounce (e.g. *people* pi:p^əl → pipɔl). The insertion of a vowel is a lenition since it eases the transition from a bilabial place of constriction /p/ to an alveolar occlusion /l/. Moreover, the CV structure is more iconic in comparison with the CCV sequence. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus expended.

6.5.8. Hiatus avoidance via linking or intrusive /r/

This process type stands for insertion of the /r/ sound intervocalically (*idea is* aɪdɪə ɪz → aɪdɪərɪz). The idea behind the insertion is to avoid the hiatus. The sequence of two vowels in the immediate vicinity seems to be ‘illegitimate’ and ‘bad’ as native speakers of English tend to break the sequence with the /r/ sound. It would be effortful to maintain a hiatus. Thus, this insertion has lenitive effects and can be classified as reduction of aerodynamic unnaturalness, which expends effort.

6.5.9. Diphthongization

This process, also referred to as breaking, inserts the schwa sound before or after a vowel within the same syllable (e.g. *me* mi: → məɪ, *fee* fi:l → frəl). The transition between sounds is prolonged. This prolonged transition results in the development of a centralized sound. The effort expended in this process consists in schwa development. Thus, the process of schwa development suppresses the natural tendency for a rapid transition from the vocalic to a consonantal or vice versa. The process can be typologized as fortition since effort is expended.

III. PROCESS GROUP: DURATION CHANGE

6.5.10. Lengthening

This process type stands for an increase in duration of a sound (e.g. *back* bæk vs. *bag* bæ:g). In particular, the vowel is prolonged in an English process which lengthens a vowel before a voiced sound. It seems that the natural tendency for a the speaker is to reduce the vowel to schwa since it requires least effort and this tendency is proven by the statistics in which the schwa is the most frequent vowel (Mitton 1992). In the case of lengthening, however, the effort takes the form of suppressing centralization. This is done in order to maintain the distinction between the voiceless/voiced consonant which follows the vowel. The process can be typologized as fortition since effort is expended.

6.5.11. Shortening

This process reduces duration of a sound (e.g. *back* bæ:k → bæk). Shortening frequently affects the vowel preceding a voiceless consonant. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided.

IV. PROCESS GROUP: ASSIMILATION

IVA. ASSIMILATION OF PLACE

6.5.12. Assimilation of stops and nasals

This process type stands for adaptation of speech sounds towards preceding or following sounds so that the feature of the neighboring sound is spread (e.g. labiality is spread *ten men* ten men → tem men, or velarization is spread *hard case* ha:rd keɪs → hark keɪs). As the result of assimilation, one place of articulation (bilabial, velar) is required for segments of the same manner (nasals, stops). The process can be typologized as lenition since gestural complexity is reduced, effort is thus reduced.

6.5.13. Palatalization

This process, also referred to as Yod coalescence, is a special case of assimilation (e.g. *as you* æz ju: → æʒju). Palatalization is distinct from other assimilatory processes due to its specific, prescribed phonetic environment: an alveolar sound must find itself in the vicinity of a palatal approximant. As a special instance of assimilation, it could belong to the complexity reduction criterion. On a closer inspection, however, palatalization operates along the natural, aerodynamic type of lenition because /ʒ/ is halfway between the alveolar ridge /z/ and the hard palate /j/. Therefore, palatalization reduces both gestural complexity and aerodynamic unnaturalness. Effort is reduced first and then expended.

IVB. ASSIMILATION OF VOICING

6.5.14. Voicing

This process involves a change from a fortis sound (a voiceless stop) to a lenis one (a voiced stop, e.g. p → b). The change reduces the amount of energy, connected with the force of articulation. /b/ requires less energy than /p/ due to the more vigorous resistance offered by the narrowed configuration of the vocal folds to the current of air. This interpretation allows to view voicing in terms of lenition understood as energy reduction (from a stronger to a weaker sound), effort being reduced.

6.5.15. Devoicing of final obstruents

This process involves a change from a weaker sound (a voiced stop) to a sound involving greater articulatory strength (a voiceless stop, e.g. b → p). There has been an animated discussion around the issue of final obstruent devoicing in English (Slowiaczek – Dinnsen 1985, Manaster Ramer 1996, Port – O'Dell 1985). Under the provision that the casual, normal English pronunciation of a voiced obstruent involves devoicing, the phrase final context makes it a suitable candidate for lenition. At the end of a phrase, the assimilation of voicing to silence occurs as the phrase is followed by an

intonational pause. Thus, reduction of gestural complexity facilitates the adaptation to silence which follows. Word final obstruent devoicing can be also classified as lenition (cf. section 6.4.2.). This process possesses both the features of lenition understood as natural outcome of aerodynamics (adaptation to silence in the phrase final context) and as reduction of complexity (voiceless stops, unlike voiced ones, do not require any action of the vocal folds/voicing). The process can be typologized as lenition since gestural complexity is reduced, effort is thus reduced.

V. PROCESS GROUP: SUBSTITUTION

6.5.16. Debuccalization

This process stands for reduction of an oral consonant to a laryngeal one (e.g. $t - \text{ʔ}$, $s \rightarrow h$). In comparison with the oral consonants, the laryngeal sounds are less complex because no velum movements are required and they are voiceless in principle. Besides, the sounds produced in the laryngeal area involve only a single contraction of the vocal folds. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided.

6.5.17. Flapping

This process stands for reduction of a stop to a flap (e.g. *dirty* $d\text{ɜ}rti \rightarrow d\text{ɜ}\text{ɹ}i$). Stops involve a ballistic movement which can be prolonged, whereas flaps involve more dynamic, ‘hit-and-run’ action of the tongue. Flaps are momentary, transient and perform a transitory movement from the starting to the finishing position. On aerodynamic grounds, flaps are easier to produce since they involve no controlled complete closure. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided.

VA. STRENGTHENING

6.5.18. Stopping

This process type stands for substitution of the more sonorous sounds with the less sonorous ones. Stopping is a change from fricative/affricate to stop, e.g. *fee* fi: → ti:. The former process is viewed as a natural outcome of aerodynamics and motorics. The complete closure requires less effort than forming a narrowed passage (Boersma 1990). Thus, the precision required for a fricative is reduced. The process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided.

6.5.19. Aspiration

This process type stands for substitution of the more sonorous sounds with the less sonorous ones (e.g. *ten* ten → t^hen). Aspiration is a fortition due to the fact that the delay in VOT is induced in an effortful way. The air pressure which is built up has a natural tendency to be released quickly since the internal and external pressures tend to equalize. Thus, the natural tendency is suppressed for emphatic purposes. The process can be typologized as fortition since effort is expended.

VB. WEAKENING

6.5.20. Spirantization

This process type stands for substitution of the less sonorous sounds with the more sonorous ones. Spirantization reduces a stop or an affricate to a fricative or an approximant, e.g. *tee* ti: → fi:). It is viewed as fortition because it is the suppression of a natural outcome of aerodynamics and motorics. Fricatives are more precise than stops due to the air passage requirements. In the linguistic literature, the ballistic movements are reported to be easier in production than controlled narrowing which has to result in a turbulent air stream (source unknown, in Boersma 1990, cf. section 1.7). The process can be typologized as fortition since effort is expended.

6.5.21. Gliding

This process type stands for substitution of the less sonorous sounds with the more sonorous ones. Gliding changes a liquid into a glide, e.g. *milk* mlk → mwk. Glides require less effort than lateral approximants. In the case of liquids, a certain degree of obstruction must be present and therefore, controlled. Glides are free from that requirement, and resemble vowels in this respect. The facts of language acquisition demonstrate that vowel production precedes that of consonants which implies that the difficulty lies in the obstruction. Since glides, unlike liquids, involve no obstruction, the process can be typologized as lenition since aerodynamic unnaturalness is reduced, effort is thus avoided.

Table 11: The revised typology of phonological processes under three lenition criteria.

Fortition	Lenition		
	Energy reduction	complexity reduction	Aerodynamic reduction unnaturalness
Aspiration	Voicing	Voicing	Epenthesis
Diphthongization		Monophthongization	Vowel insertion
Lengthening		Segment deletion	Stopping
Fricativization		Assimilation of stops and nasals	Shortening
		Degemination	Gliding
		Smoothing	Centralization
		Devoicing	Hiatus avoidance via linking or intrusive /r/
		Palatalization	Smoothing
		Centralization	Palatalization
			Debuccalization
			Flapping

Where a process appears under more than one criterion, it means that more than one criterion were evident in its operation and result.

6.6. Revision of process typology under effort management

The thesis proposes to resolve the issue of lenition/fortition typology in terms effort management (introduced in the subsection 4.2.1.). Prior to typology revision, effort management must be discussed. Effort management is functional (form follows function) but functionalism, like in NP, is modified (i.e. multifunctionalism – one from is served by many functions, the speaker is at liberty to assign functions to forms, or realize a desired result via different processes). Effort management means an optimal management of articulatory effort made by the speaker, or an aware choice of strategy (processes) in order to achieve an optimal result (communication).⁴⁹ *Optimal management* means that articulatory effort is expended in an effective way, with a view of lessening the workload in the long run, even if seemingly the speaker increases his workload. In terms of effort, a distinction must be made between effort cost and result which are often considered equal. This is the case of epenthesis in which the speaker inserts a sound but gets rid of aerodynamic unnaturalness. Another example of *optimal management* comes from language acquisition. According to Stampe (1969 [1979]), in the earliest period of babbling children try out the most radical, extreme processes because they are inborn and tested out simultaneously. Under effort management, children apply initially the most extreme processes at the expense of greater effort for a good cause: although the processes are difficult for children, they are easy at the same time because they result in the greatest contrast. In the motor speech theory (Liberman – Mattingly 1985), a child establishes a perceptual feedback as it makes a link between the movements of the vocal tract (action) and produced sounds (reaction). Thus, if a child experiences various sounds and their sequences via simultaneous application of extreme processes, the task is only seemingly difficult as the child eventually benefits from it in terms of establishing perceptual feedback and experiencing the differences between processes. *Optimal management* implies easing articulation despite seeming difficulties.

Effort made by the speaker implies that the roles of speaker and listener are not equally important since it is speaker who must make articulation easier for himself but also for the listener. It is commonly believed that in the communication act the speaker performs the function of the listener at the same time. In the course of speaking,

⁴⁹ Although the word *choice* usually appears in a collocation *a conscious choice*, the adjective *aware* is used here on purpose due to the linguistic implications of the word *conscious*. Cf. page 169.

however, the speaker has no time and intention to make his articulation listener – friendly.⁵⁰ Rather, the speaker makes an estimate of the listener’s possible response. Effort management reflects both the interests of the speaker and the needs of the listener, unlike the current approaches which view lenition as a mechanical reduction of complexity for the sake of the speaker.

Aware choice must be understood with caution. Awareness is used here in the sense of “the ability to think about and reflect upon the nature and functions of language” (James – Garrett 1991) and “the ability to focus attention on language as an object in itself or to think abstractly about language, and, consequently, to play with or manipulate language” (Jessner 2006). Awareness is, for instance, manifested in numerous studies on motherese, where mothers can reflect upon the specific properties of their speech (e.g. diminutives) addressed to children. The choice might be aware in the case of a difficult word structure, e.g. even before pronouncing the word *parallel* (whose phonetic structure is complicated due to vicinity of two liquids, one central, the other lateral) awareness is increased by the difficult structure. The choice might also be aware or the case of a communicative situation, e.g. the speaker’s awareness increases when s/he addresses his speech to a foreigner, or when speaker repeats a word). The choice might not be aware in the case of processes, in fact, the speaker is assumed not to be aware of assimilation.

An optimal result means communication. Speaker must code his/her intention in such a way that listener understands them (perceptibility) but at the same time, speaker must make it easy for him/herself (pronounceability). In terms of communicative situation, speaker manages his/her effort by increasing it (hyperarticulation in introductory token of a word) or decreasing (hypoarticulation in subsequent tokens of the word, cf. Chapter Four, 4.4.2, and 4.5.). Under effort management, speaker makes an estimate of listener’s needs and responses. Effort management also makes use of Dressler’s (1985a) idea of foregrounding and backgrounding. Effort management means that sounds or sound sequences are obscured or clarified, depending not on the total effort cost, but the result.

⁵⁰ Even if the simultaneity of speaking and listening is assumed, the speaker must give priority to speaking.

The operation of effort management can be illustrated with an example, concerning the choice between voicing and devoicing. In the example, the two realizations of /g/ are compared. The speaker has the goal s/he wants to achieve and depending on that goal, e.g. voices or devoiced a sound. The choice between voicing and devoicing is, in fact, motivated by the context. Intervocally, it is better to lenite and voice: e.g. in Polish, final obstruents are generally devoiced but in the phrase *Bóg Ojciec* is /g/, which should be devoiced to /k/ in isolation, is fully voiced, whereas before a voiceless obstruent, lenition is achieved via devoicing (e.g. in *Bóg który* /g/ is devoiced to /k/).

As far as typology revision is concerned, so far the processes have been reviewed under three criteria of lenition and the result of a process (Table 11). The criteria themselves, however, cannot serve as explanatory tools in terms of process application. As NP posits, higher principles govern preferences, which, in turn, manifest themselves as concrete, specific processes in a given language. Effort management as an overarching principle governs the application of phonological processes. Revision of processes has also taken into account how effort is managed. Thus, effort management resulted from ordering lenition according to the three criteria. From subsections 6.5.1 to 6.5.21 a conclusion can be drawn that effort management subsumes lenition and fortition under effort expenditure (fortitions and lenition criteria reducing complexity and energy), effort avoidance (lenition reducing aerodynamic unnaturalness). Thus, effort as an overarching principle of process typology is managed in three ways:⁵¹

- Avoidance - effort is avoided by substitution of more difficult sounds or sound sequences for the easier ones
- Reduction - effort is reduced, a difficult sound is deleted
- Expenditure - effort is expended and a sound can be inserted or effort is expended in the sense that the natural tendency to lenite is suppressed (fortition, which arises under effort management)

Fortition is subsumed under effort expenditure which derives from motivation for lenition. Nevertheless, effort management appears to be an overarching principle in

⁵¹ Bybee (2001) claims that phonology and sound patterns result from three natural processes (assimilation, reduction and mergers). These processes partly overlap with effort management as effort may be managed in three ways: reduced (reduction lenition processes), expended (insertion lenition processes and fortition processes) and avoided (substitution and assimilation lenition processes).

lenition/fortition definition and typology. If effort is managed (avoided and reduced), it results in lenition. If effort is managed and expended, it results in lenition and fortition. Having analyzed the phonological processes from Table 10 by means of effort management, the present thesis contributes to process typology and arranges them according to lenition criteria in Table 11 and according to effort management in Table 12.

Table 12. The revised typology of phonological processes under effort management.

effort reduction			effort avoidance			effort expenditure			
lenition									fortition
Energy reduction	Complexity reduction	Aerodynamic unnaturalness reduction	Energy reduction	Complexity reduction	Aerodynamic unnaturalness reduction	Energy reduction	Complexity reduction	Aerodynamic unnaturalness reduction	
	segment deletion	centralization	voicing	centralization	centralization			epenthesis	lengthening*
	degemination	palatalization			shortening*			palatalization	aspiration
	smoothing				stopping			hiatus avoidance via linking or intrusive /r/	spirantization*
	assimilation of stops and nasals				debuccalization				diphthongization*
	devoicing of final obstruents				flapping				
	palatalization				gliding				
	monophthongization*								

Where a process appears under more than one criterion, it means that more than one criterion were evident in its operation and result.

The following five processes: monophthongization, diphthongization, lengthening, shortening and spirantization were asterisked. Its placement is considered with respect to the first four processes first, whereas the final one is considered separately.

In revising the process typology, the argument for a certain criterion was that it *eases* pronunciation. It must be specified here that although the present thesis uses the notion of *phonological processes* in the sense of Donegan and Stampe (1979), one important reservation can be made. This reservation concerns the notion of *ease of articulation*, which is evoked by Donegan and Stampe (1979) in explaining phonological processes. The thesis employs the abovementioned authors' understanding of processes; ease of articulation, however, is treated in a different way. The present

thesis adopts effort management in the understanding of ease of articulation. According to effort management, the speaker must mediate between his/her ease and the listener's ease. In general, the notion is surrounded by numerous controversies and has become a fiercely debated issue in linguistics, especially when it is invoked in the context of language change. For instance, for Trask (1996) the notion of *ease* lacks higher motivation: “[t]he degree to which speech sounds, sequences of speech sounds and contrasts between speech sounds may be readily produced by speakers, sometimes cited as a factor on phonological change. This notion underlies the notion of natural processes” (Trask 1996: 126). Ohala (1984) called it a *wild card* when it is used in the context of a historical development of language:

Fischer-Jørgensen: (to Ohala) why - in your paper - do you reject explanations of sound change based on "ease of articulation"? Ohala: It is a matter of research strategy. We know so little about the 'effort' involved in articulating sounds that it is a notion that is too easy to invoke and, frankly, it is a notion that has often been abused. We should exhaust the explanatory principles that are known and testable before using this 'wild card' (Ohala 1984).⁵²

Gussmann (Edmund Gussmann, p.c.) referred to *ease* as to a *moronic term*. The above criticism is not unfounded since the wording of the concept is indeed unfortunate. *Ease* may mean both speaker and the listener. For instance, the cluster /pw/ is easy for the speaker due to its labiality but difficult for the listener as it introduces no contrast. On the other hand, the sequence /pa/ is easy for the listener but considered difficult for the speaker. It is a common belief in linguistics that maintaining articulators in the same place of articulation (steady state) such as in /pw/ is easier for the speaker. “According to well-known study by Janson (1986: 193), favored combinations are those in which the articulators do not have to make extensive movements from the consonant gesture to the vowel gesture” (Dziubalska-Kończak 2002a: 98). Easiness of steady states for the speaker is also supported by the proximity law (Dziubalska-Kończak 2002a). The assumption that *ease of articulation* serves the speaker can be challenged. It may be argued that maintaining articulators in the same place of articulation is more effort-costly than changing their configuration. Intuitively, maintaining a configuration for a longer time involves more effort in terms of precision (the articulators cannot move and must stay in the same position) and energy (maintaining must be forced upon articulators). If one compares the vocal tract to a motor system and articulators to motors, then indeed

⁵² This quotation has been provided by Zofia Malisz (p.c.).

it makes more sense to call maintaining steady states difficult. Thus, the thesis argues that steady states are difficult not only for the listener, but also for the speaker. This might be the case of the clusters /wu/ and /ji/ which are difficult both for perception and production. In similar vein, diphthongs are better than long vowels. The claim that steady states are difficult for the speaker is postulated in the absence of any convincing evidence for their *easiness*. Moreover, it derives an argument from phonostylistic processes. The phrase *ten men* is assimilated so that the place and manner of articulation is the same (bilabial nasal) and the phrase becomes *tem men*. Note that in casual speech, however, double bilabial nasal is degeminated. This can be linked to the claim that steady states (*mm*) are more difficult for the speaker. Perhaps s/he finds it not easy to maintain articulators in the same place of articulation, thus shortens the geminate. Therefore, asterisks were placed next to the opposite processes such as lengthening/shortening and monophthongization/diphthongization. Unfortunately, the discussion which action: changing or maintaining articulation is easier or more difficult cannot be properly conducted since there are no relevant descriptions of this kind in the linguistic literature. It must be admitted that typologization of lengthening/shortening and monophthongization/diphthongization presents some difficulties in the absence of relevant research.

The above discussion clarifies the use of asterisk with reference to monophthongization, diphthongization, lengthening and shortening. Sprantization, however, presents a different typology problem than the issue of maintaining/changing a state. In Table 12, it was typologized as fortition on the grounds that the process expends effort and violates the aerodynamic unnaturalness as it is more natural in terms of the vocal tract dynamics to produce a stop. If one considers inertia, spirantization can be classified as lenition deriving from reduction aerodynamic unnaturalness. If fricativization is considered intervocalically, e.g. the intervocalic stop in *baba* can be spirantized to adjust the opening for vowels (Jaworski 2007) and sprantization here gains the status of easing articulation (lenition). Thus, one could suggest that lenition criteria do not exhaust the explanation possibilities in explaining phonological processes typology. If in one context spirantization is considered as lenition and as fortition in another, it is tempting to introduce context as another parameter of phonological processes. Unfortunately, reduction of gestural complexity, unlike reduction of aerodynamical unnaturalness, is not subject to context (cf. 6.4., Figure 40). Introduction

of context into all lenition criteria, however, would make all attempts of typology futile as with the context in all cases, everything becomes relative. Effort management, regardless of the context, serves as the overarching principle which governs process typology.

6.7. Teleological perspective

Typology based on effort management can be further explained by means of goal-oriented behavior which achieves fortitive or lenitive results regardless of the operational procedure. For instance, devoicing can be understood as complexity or gestural reduction but the goal (lenition) is achieved in one of the two ways. This also implies that effort management-based process typology is teleological. Therefore, the present section discusses teleology in greater detail.

Itkonen (1978a, 1978b, 1983) made a claim that the least effort principle motivates teleology. This claim carries great significance and allows adopting teleological explanation into process typology under effort management. Effort management as the theory behind the process typology receives therefore a deeper conceptualization. Since language has a communicative function, the phonological processes serve the ultimate goal: to be understood. Regardless of the strategy, the selection of a process is result-oriented (i.e. teleological). In emphatic speech, when lenition is likely to be penalized, the speaker consciously suppresses lenition and succeeds in communication in this particular speech style. The selection aims at the final result, rather than at strengthening the sounds for the art's sake. In this way, teleology serves as a powerful piece of evidence that process typology should not have mechanical or causal criteria. In the case of fortitions, sound strengthening is not accidental, but motivated by the overall result. Thus, teleology validates effort management as a possible candidate for lenition/fortition typology instead of the current weakening/strengthening ones. Moreover, implementation of teleology helps to avoid the problems of the traditional approaches to process typology which were connected with defining lenitions/fortitions.

Therefore, the present thesis employs teleology as the explanation type for the proposed typology of fortitions/lenitions. Teleology is a philosophical doctrine which stipulates the existence of final causes, purposes and design. The doctrine explains phenomena by their ends or purposes⁵³. A *Dictionary of Philosophy* defines teleology in the following way: “[t]he theory of purposiveness in nature: characteristically, certain phenomena seem to be best explained not by means of prior causes, but ends or aims, intentions or purposes. Teleological explanations seem typical of living or organic things – plants, animas, people etc.” (A *Dictionary of Philosophy* 1979: 325). Under teleological perspective, purpose and design are a part of nature or they are apparent in nature. Therefore, teleology stands in opposition to the mechanism theory which assumes that all events may be explained by mechanical principles of causation. Therefore, phenomena are governed not only by mechanical forces but they also strive to achieve the goals of self-realization or find purpose through their design. Numerous disciplines of science, such as natural sciences and theology, employ the doctrine. The classic example of teleological arguments identifies the existence of the universe with God’s will. The arguments of this sort bear the name of *argument from design*.⁵⁴ They also account for the construction of this world and all living things in a telic, purposeful rather than a chaotic, incidental manner. These arguments imply that the designer must be a rational being (God) since the design is purposeful (Gasson 1973, Rescher 1986). This implication leads to the development of the *creationist perspective* which stipulates that this world had been created by God.

Alternative hypotheses explaining the world’s origins exclude the divine intervention. Darwin (1859 [1964]) advanced the evolutionist prediction that species develop by natural selection, which, in turn, also applies the concept of purpose. The *survival of the fittest* has ushered in a new quality of the notion of purpose: for instance, a lion has claws so it can hunt and kill its prey. The fact that the lion has claws gains it an upper hand in the selection process. The lion occupies a higher rank in the food chain as a predator, which enhances its chances of survival. Modern physics and astronomy account for the world’s origins in terms of the *Big Bang theory* (Lemaître 1927). The big bang was originally an explanation of the fact that distant galaxies are traveling away from the earth at great speeds (Hubble 1936). According to the theory, a cosmic

⁵³ From Greek: *téleos* – “perfect, complete” and *télos* – “end, result” (Kopalinski 1989).

⁵⁴ The argument from design has frequently come under criticism since it lacks a scientific element. Any hypotheses of the sort “X causes Y” without sufficient proof fall within the category of illegitimate teleologies, or are at peril of being classified as a logical fallacy.

explosion of a primeval atom began the universe, which emerged from a dense and hot state. It is calculated that the initial explosion took place about 13.7 billion years ago. Nevertheless, teleology fails to receive an explicit recognition as an explanatory strategy:

[T]eleology as such is an intimidating topic, due to the multiplicity of – often vague – meanings associated with the notion, to the many conflicting interpretations of its ontological and epistemological status and to the common prejudice against teleological explanation, which results from the unwarranted conviction that it is a metaphysical and not a scientific concept (Adamska-Sałaciak 1986: 3 - 4).

The term *teleology* was coined as late as 18th century with reference to the study of final causes in nature. The origins of teleology can be traced back to Aristotle (350 B.C.E. [1962]) who formulated nature laws. According to the nature laws, each phenomenon results from a unity between form and substance. Form determines class membership whereas substance determines uniqueness. For Aristotle (350 B.C.E. [1962]), the intended function of a thing was its final cause. He also distinguished between material, formal, efficient and final aspects of causation. The final causation has primacy over the other aspects since the final one is teleological. Aristotelian views stipulated an overlap between the formal and final causation: the growth of a rosebud is its development into a fully developed rose. Therefore, the process is complete: the existence of a rosebud (formal causation) serves a rose (final causation). The ideas developed by Aristotle (350 B.C.E. [1962]) commenced the dichotomy between causal and teleological explanation.⁵⁵ Teleological explanation lays emphasis on the purpose of existence (final cause), whereas the causal one is concerned primarily with laws governing the development of phenomena out of each other (efficient causes). The latter one, represented by Galilean type of explanation, has proved a huge success in the exact sciences (Adamska-Sałaciak 1986).

The causal explanation, also referred to as *reductionism*, rejects teleological explanation on the basis of the argument that teleology is not an explanation *sui generis*. Reductionism exists in its weak and strong versions. The weak version reduces teleological explanation to the causal ones, differentiation between different degrees

⁵⁵ There is also the third type of explanation in the scientific tradition except for the causal and teleological ones. The third type, influenced by Newtonian physics, explains phenomena via mechanical laws. Kant (1790 [1964]) unites the mechanical and teleological explanation types. Under his approach, human mind responds mechanically the teleological principle, as if it was preprogrammed. This type of explanation, however, fails to incite much interest among the explanation theories.

(elliptic, partial and sketchy explanation). In its strong version, reductionism employs the three doctrines: the good consequence doctrine, the explanation/etiology doctrine and the goal doctrine. They differ in their approaches to function. The good consequence doctrine focuses on “process taking place in living organisms” (Adamska–Sałaciak 1986: 50). It stipulates that phenomena occur because they are good, i.e. something is a function only and only if it is conferring some good. For instance, the function of heart is not producing audible heart beats, but ensuring blood circulation. Blood circulation is good in the sense of keeping the organism alive. Unfortunately, goodness as a requirement of function is difficult to capture. Thus, the good consequence cannot be effective in a methodological way. The explanation/etiology doctrine studies human and animal behavior. According to the doctrine, actions and behaviors possess certain identifiable purposes but should not be explained through the purposes themselves. This line of thinking reduces teleology to a mere type of causal explanation. Its major weakness consists in focusing on etiology itself instead of its purposes: the doctrine holds that “teleological statements do not normally count as explanations if by explanations we mean an account of how the sort of thing we are trying to explain comes to be as it is” (Adamska–Sałaciak 1986: 43). The goal doctrine, modeled on cybernetics, assumes that a system is goal-oriented. The system is characterized by plasticity. Plasticity denotes the possibility of accomplishing goals via a set of alternative routes or methods. Persistence, the other feature of the system, ensures that a state or a property continues to manifest itself and ensures such direction of a system development which allows the system to accomplish its goals.

Regardless of the advancements of reductionism, teleological explanation is “fully legitimate methodologically” (Adamska–Sałaciak 1986: 79). This is so due to the fact that it bears a strong resemblance to rational and pattern explanation. It can be also compared to functional explanation.⁵⁶ The primacy of the teleological explanation over the causal ones is illustrated by Adamska–Sałaciak (1986) with special reference to the phenomenon of diachronic linguistic change. She derived her argument from the social character of language: “all change is teleological in the most general sense (i.e. in the sense of ensuring the continuity of language)” (Adamska–Sałaciak 1986: 277). Therefore, the individual, psychological element in language which is stipulated by the

⁵⁶ Functional explanation is not a logical derivative of the teleological one. The former one employs activities which contribute to an end (function), whereas the latter one employs the activities which contribute to the end (purpose) (Adamska – Sałaciak 1986).

causal explanation, must yield to the primacy of the sociolinguistically motivated goals. For instance, mergers can be accounted for by the communicative (social) function of language. Coseriu (1974) supports teleological explanation of change. He proposes a finalistic approach, based on the interpretation of language as *energeia*. If language is energy, it is subject to development and strives to achieve finalistic purposes in terms of an optimal state. This finalistic theory is nothing else but a variety of teleological approach. For Itkonen (1978a, 1978b, 1983), change resulted from human actions which, in turn, conform to socially valid rules. He advocates the teleological explanation and draws a distinction between the two types of teleology in linguistic change: the short- and long-term teleology. The former one is language universal and is illustrated by the isomorphism principle (i.e. one meaning has one and only one form). This type of teleology is, among others, motivated by the principle of least effort. The latter, the short-term teleology is language specific. This teleology strives to achieve the subsystem-internal goals (e.g. a drift in terms of the switch from an OV to a VO order). The long-term teleology is reduced to the short-term one (Itkonen 1978a, 1978b, 1983).

Chapter Seven

Conclusions

Investigating the principle of least effort, the thesis discusses the following notions and concepts: the principle of least effort (Chapter One), the concept of hierarchy (Chapter Two), the concept of preferences (Chapter Two), the notions of easy and difficult sounds (Chapter Three), the concept of articulatory effort (Chapter Three), articulatory gestures (Chapter Three), phonological processes (Chapters Five and Six), typology and teleology (Chapter Six).

Although the concepts have received considerable scholarly attention in terms of an extensive literature, in investigating them the thesis makes a number of new proposals: a measure of articulatory effort via parameters, the attention hypothesis, criteria of lenition other than energy (i.e. complexity and aerodynamic unnaturalness), alternative definitions of lenition and fortition, effort management, a revised typology of phonological processes and treatment of the notion *ease of articulation* with skepticism in explaining phonological processes.

The idea that articulatory effort can be measured via parameters has already been proposed: effort as a total of energy, the presence of articulatory gestures, synchronization of gestures, precision, systemic effort, and coordination) (Boersma 1998). The thesis, however, does not pursue this particular proposal. It introduces five parameters of articulatory effort: vowel duration; vowel centralization as vowel height and vowel position; closure duration and its voicing and the degree of spirantization, all of which can be analyzed in Praat. Since all the proposed parameters provide an equal

contribution to effort measure and prove equally representative for measuring articulatory effort (the first experiment, cf. Chapter Four), the thesis selects only one of them, i.e. vowel duration (the second and third experiments, cf. Chapter Four). A possible venue for further research would be creating an integrated, holistic approach to articulatory effort measure via establishing the role of individual parameters such as those proposed by Boersma (1998) and Maddieson (p.c).

The attention hypothesis is proposed in the course of interpreting the results of the third experiment. The hypothesis postulates the cyclical nature of attention, i.e. that it is high in the beginning and in the end like in the bathtub effect. Unfortunately, the literature search provided no such parallel hypothesis in psychology; therefore, the proposal remains a tentative hypothesis. The question whether human attention (defined in terms of acoustic distinctiveness) indeed rebounds in the end of performed task calls for more research.

In investigating the principle of least effort, the thesis identifies two issues in the area of phonological processes: the issue of defining lenition/fortition and the issue of criteria for process typology. The thesis proposes to define lenition as reduction and proposes its three hierarchically organized criteria: reduction of energy, reduction of complexity and reduction of aerodynamic unnaturalness, whereas fortition is defined in the thesis as an effortful suppression of lenition. In the absence of a comprehensive definition of lenition/fortition, the new lenition criteria and a negative definition seem to be of particular significance for Natural Phonology. Claims such as the following one made by Kirchner (1998):

The Natural Phonology program of Stampe (1972) and Donegan & Stampe (1979) attacked the phonetic arbitrariness of classic Generative Phonology. Anticipating much of the orientation of the effort-based approach, Donegan & Stampe invoke the twin functional principles of ease of articulation and ease of perception; and in fact use the term "lenition" to cover all patterns motivated by the former, including articulatorily driven assimilations, such as /nb/ - [mb]. Moreover, they make a number of proposals concerning the *ordering* of lenition rules, relative to fortition rules. Unfortunately, the Natural Phonology program did not develop a restrictive, unified formal characterization of lenition processes. In the absence of a formalism capable of expressing violable conflicting principles, the functional insights of Natural Phonology remained unformalized metatheory (Kirchner 1998: 13-14).

appear to lose their validity in the light of the fact that the present thesis attempts to formulate criteria for phonological processes and their typology.

The thesis also revises the current process typology in terms of effort management. The thesis proposes to use the term *effort management* instead of *least*

effort in order to incorporate teleology into the principle in a more evident way since the latter term fails to do justice to the term's implications. The term *least effort* suggests that the workload is somehow avoided, unlike the term *effort management* which stands for an optimal expenditure of effort and a selection of a path to follow. *Effort management* is functional and means an optimal management of articulatory effort made by the speaker, or an aware choice of strategy (processes) in order to achieve an optimal result (communication). Effort management is used in the thesis to revise the current typology of phonological processes. Thus, effort as an overarching principle of process typology is managed in three ways: effort avoidance, reduction and expenditure. Some processes (monophthongization, diphthongization, lengthening and shortening), however, are difficult to handle by effort management due to the fact that there is no conclusive evidence in the literature whether maintaining or changing place or manner of articulation is easier. Thus, the related notion *ease of articulation* must be treated with skepticism. This skepticism calls for further research on the issue whether maintaining steady articulation states or changing them is more effective in terms of articulatory effort.

The two former proposals, i.e. definition of lenition/fortition and revised process typology are the thesis' contribution to the area of phonological processes. The contribution lies in the improvements on approaches to phonological processes dealt with in Chapter Six. The thesis incorporates gestures and the dynamics of the vocal tract into the definition of lenition/fortition and proposes effort management in process typology.

The adoption of effort management as a principle governing the process typology grants it the status of an overarching principle which a particular process is subject to. The implications of the assumed status appear to be far-reaching. Since effort management works for the benefit of the speaker, the speaker may behave more egoistically than is commonly assumed. Thus, the dichotomy between the speaker and the listener in terms of interests and needs appears to lose its validity, the focus instead being exclusively on the speaker's interests. This implication is in opposition to Ohala's (1993) claim that the listener is the source of all linguistic changes and that his needs/interests govern those of the speaker's. Hopefully, the thesis revives the linguistic discussion about the clear-cut division between production and perception, or the role of the speaker and the listener in the communication act.

References

- Adamska – Sałaciak, Arleta. 1986. *Teleological explanations in diachronic phonology*. [PhD dissertation, Adam Mickiewicz University].
- Allen, Timothy F. H. [nd] *A summary of the principles of hierarchy theory*. (<http://www.iss.org/hierarchy.htm>) (date of access: 6 May 2004).
- Allen, Timothy F. H. and Thomas B. Starr. 1982. *Hierarchy: perspectives for ecological complexity*. Chicago: University Chicago Press.
- Aristotle. 350 B.C.E. [1962]. *Nicomachean ethics*. (Translated by Martin Ostwald.) Indianapolis: Bobbs – Merrill.
- Audi, Robert. (ed). 1999. *Cambridge dictionary of philosophy*. Cambridge: Cambridge University Press.
- Aurelius, Marcus. 167 AD [1998]. *The meditations of Marcus Aurelius Antoninus*. (Translated by A. S. L. Farquharson). Oxford: OUP.
- Austin, John L. 1962. *How To Do Things with Words*. Oxford: Oxford University Press.
- Aylett, Matthew P. 2000. *Stochastic suprasegmentals: relationships between redundancy, prosodic structures and care of articulation in spontaneous speech*. [PhD dissertation, University of Edinburgh]. (<http://www.cogsci.ed.ac.uk/~matthew/icslp2000.ps>) (date of access: 18 May 2004).
- Bartra, Anna et al. (eds.). 1984. *Estudis grammaticals I, Publicacions del Departament de Filologia Hispanica de la Universitat Autònoma de Barcelona Serie Lingüística*. Universitat Autònoma de Barcelona: Bellaterra.
- Baudouin de Courtenay, Jan Niecisław. 1877 [1972]. *A Baudouin de Courtenay anthology. The beginnings of structural linguistics*. (Translated and edited with

- an introduction by Edward Stankiewicz.) Bloomington: Indiana University Press.
- Baudouin de Courtenay, Jan Niecislaw. 1895. *Versuch eine Theorie phonetischer Alternationen: Ein capitel aus der Psychophonetik [An attempt of phonetic alternation theory: a chapter on psychophonetics]*. Strassburg: Trübner.
- Billerey-Mosier, Roger. 2000. *Lexical effects on the phonetic realization of English segments*. [MA thesis, UCLA].
(<http://www.linguistics.ucla.edu/people/grads/billerey/billerey.htm>) (date of access: 9 Jan. 2005).
- Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: CUP.
- Boersma, Paul. 1990. *Modelling the distribution of consonant inventories*. (Paper presented at the Congress of Linguistics and Phonetics, 30 August 1990, Prague, the Czech Republic). (http://www.hum.uva.nl/paul/papers/Praag_1990.pdf) (date of access: 4 April 2004).
- Boersma, Paul. 1997. *The elements of functional phonology*. Manuscript, University of Amsterdam. ROA-173, Rutgers Optimality Archive.
(<http://rucss.rutgers.edu/roa.html>) (date of access: 10 April 2005).
- Boersma, Paul. 1998. *Functional phonology. Formalizing the interactions between articulatory and perceptual drives*. Utrecht: LOT.
- Boersma, Paul – David Weenink. 2005. Praat: doing phonetics by computer (Version 4.3.22) [Computer program]. Retrieved September 8, 2005, from <http://www.praat.org/>
- Bopp, Franz. 1863. *Kritische grammatik der Sanskrit-sprache [A critical grammar of Sanskrit]*. Berlin: Nicolai.
- Bopp, Franz. 1989 [1974]. *Analytical comparison of the Sanskrit, Greek, Latin, and Teutonic languages, shewing the original identity of their grammatical structure: new edition*. Amsterdam: John Benjamins Publishing Company.
- Van den Broecke, Marcel P.R. – Antonie Cohen (eds). 1984. *Proceedings of the tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications.
- Brogan, Albert P. 1919. "The Fundamental Value Universal", *Journal of Philosophy, Psychology, and Scientific Methods* 16: 96-104.
- Bronstein, Ilja N. – Konstantin A. Semendjajew. 1996. *Matematyka. Poradnik encyklopedyczny. [The handbook of mathematics]*. Warszawa: PWN [Polish Scientific Publishers].

- Browman, Catherine P. – Louis M. Goldstein. 1986. "Towards an Articulatory Phonology", *Phonology Yearbook* 3: 219-252.
- Browman, Catherine P. – Louis M. Goldstein. 1992. "Articulatory Phonology: an overview", *Phonetica* 49: 155-180.
- Bussmann, Hadumod. 1996. *Routledge dictionary of language and linguistics*. London and New York: Routledge.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: CUP.
- Clements, George N. 1990. "The role of the sonority cycle in core syllabification", in: John Kingston – Mary E. Beckman (eds.), 283-333.
- Chomsky, Noam. 1965. "Three models for the description of language", in: R. Duncan Luce – Robert R. Bush – Eugene Galanter (eds.), 105-124.
- Chomsky, Noam – Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Coseriu, Eugenio. 1968. "Sincronía, diacronía y tipología" [Synchrony, diachrony and typology], *Actas del XI. Congreso Intern. de lingüística y filología románicas*. Madrid: CSIC, 269-283.
- Coseriu, Eugenio. 1974. *Synchronie, Diachronie und Geschichte: das Problem des Sprachwandels [The synchrony, diachrony and history: the problem of sound change]*. (Translated by Helga Sohre.) München: W. Fink.
- Coulmas, Florian. 1992. *Language and economy*. Massachusetts/Oxford: Blackwell.
- Crystal, David. 1995. *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Curtius, Georg. 1858-1862. *Grundzüge der griechischen Etymologie [The principles of Greek etymology]*. Leipzig: Teubner.
- Dalcher, Christina Villafañá. 2006. *Consonant Weakening in Florentine Italian: An acoustic study of gradient and variable sound change*. [PhD dissertation, Georgetown University].
(<http://www.georgetown.edu/users/cmv2/Dissertation.htm>) (date of access: 27 Feb 2007)
- Darwin, Charles. 1859 [1964]. *Origin of species by means of natural selection, or the preservation of favored races in the struggle for life*. Harvard: Harvard University Press.
- A Dictionary of Philosophy*. 1979. London: Pan Books.

- Dinnsen, Daniel (ed.). 1979. *Current approaches to phonological theory*. Bloomington: IUP.
- Dinnsen, Daniel. 1980. "Phonological rules and phonetic explanations", *Journal of Linguistics* 16: 171-191.
- Docherty, Gerard J. – D. Robert Ladd (eds.). 1992. *Papers in Laboratory Phonology II: Gesture, segment, prosody*. Cambridge: CUP.
- Donegan, Patricia – David Stampe. 1979. "The study of Natural Phonology", in: Daniel A. Dinnsen, (ed.), 126-173.
- Donegan, Patricia. 1985. *On the natural phonology of vowels*. New York: Garland.
- Donegan, Patricia. 2002. "Phonological processes and phonetic rules", in: Katarzyna Dziubalska-Kołaczyk – Jarosław Weckwerth (eds.). 57-81.
- Dressler, Wolfgang U. – Wolfgang Meid (eds.). 1978. *Proceedings of the 12th International Congress of Linguists*. Innsbruck: University Innsbruck.
- Dressler, Wolfgang U. 1985a. *Explaining Natural Phonology. Phonology Yearbook* 1. 29-50.
- Dressler, Wolfgang U. 1985b. *Morphonology: the dynamics of derivation*. Ann Arbor: Karoma Press.
- Dressler, Wolfgang U. 1997. *On semiotic theory of preferences: Proceedings of the international colloquium on language and Peircean sign theory*. New York: Bergham Books.
- Dziubalska-Kołaczyk, Katarzyna (ed.). 2001. *Constraints and preferences*. Trends in Linguistics. Studies and Monographs 134. Berlin: Mouton de Gruyter.
- Dziubalska-Kołaczyk, Katarzyna. 2001. "Phonotactic constraints are preferences", in: Katarzyna Dziubalska-Kołaczyk (ed.), 69-100.
- Dziubalska-Kołaczyk, Katarzyna. 2002a. *Beats-and-Binding Phonology*. Frankfurt: Peter Lang.
- Dziubalska-Kołaczyk, Katarzyna – Jarosław Weckwerth (eds.) 2002b. *Future challenges for Natural Linguistics*. Munich: Lincom.
- Dziubalska-Kołaczyk, Katarzyna. 2002b. "Challenges for Natural Linguistics in the twenty first century: a personal view", in: Katarzyna Dziubalska-Kołaczyk – Jarosław Weckwerth (eds.), 103-128.
- Dziubalska-Kołaczyk, Katarzyna. 2003. "Speech is in the ear of the listener: Some remarks on the acquisition of second language sounds", in: Kimberli Hales – Angela Terveen (eds.), 81-92.

- Dziubalska-Kořaczyk, Katarzyna. 2006. "Modern Natural Phonology. The theory for the future", in: Jacek Fisiak (ed.), 1-10.
- Edgeworth, Francis Ysidro. 1881 [1967]. *Mathematical psychics*. New York: Augustus M. Kelley.
- Ernestus, Mirjam. 2000. Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface. LOT: Utrecht.
- Ernst, Peter (ed.). 1999. *Einführung in die synchrone Sprachwissenschaft [An Introduction to Synchronic Linguistics]*. Vienna: Edition Praesens (2nd edition).
- Ethnologue. 2005. *Bibliography database*.
(http://www.ethnologue.com/show_subject.asp?code=TON) (date of access: 27 April 2005).
- Estoup, Jean-Baptiste. 1916. *Gammes stenographiques [The stenographic sings]*. Paris: Institut Stenographique de France.
- Fenk, August – Gertraud Fenk-Oczlon. 1993. "Menzerath's Law and the constant flow of linguistic information", in: Reinhard Köhler – Burghard B. Rieger (eds.), 11-31.
- Ferguson, Charles, A. 1978. "Phonological processes", in: Joseph H. Greenberg (ed), 405-445.
- Figure/ground. [nd] (http://www.psychologie.tu-dresden.de/i1/kaw/diverses%20Material/www.illusionworks.com/html/figure_ground.html) (date of access: 12 June 2007).
- Fisiak, Jacek (ed.). 2006. *English language, literature and culture. Selected papers from the 13th PASE conference, Poznan 2004*. Poznań: Uni-Druk.
- Flemming, Edward. 1995. *Articulatory features in phonology*. [PhD dissertation, California University].
- Foley, James. 1977. *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.
- Fried, Vilem (ed.). 1982. *Selected papers in structural linguistics. Contributions to English and general linguistics written in the years 1928-1978*. Berlin: Mouton Publishers.
- Frigeni, Chiara. 2006. Review of "Lenition and contrast" by Naomi Gurevich. *The LINGUIST List*. (<http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/pubs/reviews/get-review.cfm?SubID=66328>) (date of access: 23 Jan 2006).

- Frisch, Stefan A. 1998. Review of “Phasing and recoverability” by Daniel Silverman. *The LINGUIST List*. (<http://www.linguistlist.org/issues/9/9-1588.html>) (date of access: 3 March 2003).
- Fujimura, Osamu. 1981. “Temporal organization of articulatory movement as a multidimensional phrasal structure”, *Phonetica* 38: 66-83.
- Gasson, Peter. 1973. *Theory of design*. London: Harper and Row.
- Gafos, Adamantios I. 1999. *The articulatory basis of locality in phonology*. New York: Garland Publishers. (<http://homepages.nyu.edu/~ag63/>) (date of access: 9 Jan. 2005).
- Gengerelli, Joseph Anthony. 1930. “The principle of maxima and minima in animal learning”, *Journal of comparative psychology* 11: 193-236.
- Gimson, Arnold C. 1995. *An introduction to the pronunciation of English*. 5th edition revised by Alan Cruttendon. London: Edward Arnold.
- Goman, Roderick David. 1979. *Consonants in Natural Phonology*. [PhD dissertation, the Ohio State University].
- Grammont, Maurice. 1901. “Onomatopées et mots expressifs” [Onomatopoeias and expressive words], *Trentenaire de la société pour l'étude des langues romanes*, 261-322, Montpellier.
- Greenberg, Joseph H. 1968. *Anthropological linguistics*. New York: Random House.
- Greenberg, Joseph H. (ed.). 1978. *Universals of human language*, Volume 2: Phonology. Stanford: Stanford University Press.
- Grimm, Jacob. 1918. *Deutsche grammatik [The German grammar]*. Göttingen: Dietrich.
- Grunwell, Pamela. 1982. *Clinical phonology*. Guildford, London, Oxford, Worcester: Billing and Sons Limited.
- Gurevich, Naomi. 2004. *Lenition and contrast*. New York and London: Routledge.
- Hales, Kimberli– Angela Terveen (eds.). 2003. *Selected papers from the Sixth College-wide Conference for Students in Languages, Linguistics and Literature 2002*. Honolulu: University of Hawai‘i at Mānoa.
- Hansson, Sven Ove. 1984. *Decision theory. A Brief Introduction. [Incomplete book manuscript]*. (<http://www.infra.kth.se/~soh/decisiontheory.pdf>) (date of access: 15 February 2006).
- Hardcastle, William. J. – Alain Marchal (eds.). 1990. *Speech production and speech modeling*. Dordrecht: Kluwer Academic Publishers.

- Harris, James. 1984. "La espirantización en castellano y su representación fonológica Autosegmental" [Spirantization in the Spanish language and its representation in autosegmental phonology], in: Anna Bartra et al. (eds.), 149-167. (<http://web.mit.edu/fll/www/publications/JamesHarris.html>) (date of access: 15 April 2006).
- The Harvard School of Public Health. 2004. *Food pyramids. What Should You Really Eat?* (<http://www.hsph.harvard.edu/nutritionsource/pyramids.html>) (date of access: 10 April 2005).
- Hayes, Bruce – Robert Kirchner – Donca Steriade (eds.). 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Hofstadter, Douglas R. 1979. *Godel, Escher, Bach: an eternal golden braid*. New York: Basic Books.
- Hoopes, James (ed.). 1991. *Peirce on Signs: Writings on semiotics*. Chapel Hill: University of North Carolina Press
- Horn, Laurence R. 1984. "Toward a new taxonomy for pragmatic inference: Q – based and R – based implicature", in: Deborah Schiffrin, 12-42.
- Hoyt, Donald F – Richard C. Taylor. 1981. "Gait and the energetics of locomotion in horses", *Nature* 292: 239.
- Hubble, Edwin. 1936. *The Realm of the Nebulae*. New Haven: Yale University Press.
- Hull, Clark Leonard. 1943. *Principles of behavior*. New York: Appleton-Century.
- Hyman, Larry. M. 1975. *Phonology: theory and analysis*. New York: Holt, Rinehart&Winston.
- Hyman, Larry. M. – Charles N. Li (eds.). 1988. *Language, Speech and Mind*. London: Routledge.
- Itkonen, Esa. 1978a. *Grammatical theory and metascience*. Amsterdam: John Benjamins.
- Itkonen, Esa. 1978b. "Linguistics: Nonempirical and empirical", in: Wolfgang U. Dressler – Wolfgang Meid (eds.), 157-158.
- Itkonen, Esa. 1983. *Causality in linguistic theory*. London: Croom Helm.
- Jacob, Evelyn. 1990. "Alternative approaches for studying naturally occurring human behavior and thought in special education research", *The Journal of Special Education*, 24,(2), 195-211.
- Jakobson, Roman. 1941 [1968]. *Child language, aphasia and phonological universals*. The Hague: Mouton.

- James, Carl – Peter Garrett (eds.). 1991. *Language awareness in the classroom*. London: Longman.
- Jaworski, Sylwester. 2007. *Lenition processes in English and other languages: A hierarchy of susceptibility to inertia*. [Unpublished PhD dissertation, Adam Mickiewicz University].
- Jespersen, Otto. 1921 [1964]. *Language: its nature, development and origin*. New York: Norton.
- Jessner, Ulrike. 2006. *Linguistic Awareness in Multilinguals*. Edinburgh: Edinburgh University Press.
- Jones, Charles (ed). 1993. *Historical linguistics: problems and perspectives*. London/New York: Longman.
- Jun, Jongho. 2004. "Place assimilation", in: Bruce Hayes – Robert Kirchner – Donca Steriade (eds.), 58-87.
- Kant, Immanuel. 1790 [1964]. *Kritik der Urteilstkraft [Critique of judgment]*. (Translated by J.C. Meredith.) Oxford: Clarendon Press.
- Kellerman, Günter – Michael D. Morrissey (eds.). 1992. *Diachrony within synchrony. language history and cognition*. [Duisburger Arbeiten Zur Sprach-und Kulturwissenschaft 14]. Frankfurt an Main: Peter Lang Verlag.
- Kelso, Scott. J.A. 1984. "Phase transitions and critical behavior in human bimanual coordination", *American Journal of Physiology: Regulatory, Integrative and Comparative* 15: R1000-R1004.
- Kenstowicz, Michael – Charles W. Kisseberth (eds.). 1973. *Issues in Phonological Theory*. The Hague: Mouton.
- Kingston, John – Mary E. Beckman (eds.). 1990. *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.
- Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. [PhD dissertation, UCLA] (<http://www.ruccs.rutgers.edu/ROA>) (date of access: 30 May 2004).
- Kiritani, Shigeru. 1986. "X-Ray microbeam method for measurement of articulatory dynamics techniques and results", *Speech Communication* 5: 119-140. North Holland.
- Kopaliński, Władysław. *Słownik wyrazów obcych i zwrotów obcojęzycznych [A dictionary of foreign words]*. 1989. Warszawa: Wiedza Powszechna.

- Köhler, Reinhard – Burghard B. Rieger (eds.). 1993. *Contributions to quantitative linguistics*. The Netherlands: Kluwer Academic Publishers.
- Kreidler, Charles W (ed.). 2001. *Phonology: critical concepts. Volume I: A variety of viewpoints in phonology*. London and New York: Routledge.
- Krug, Manfred. 1998. "String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change", *Journal of English Linguistics* 26: 286-320.
- Kruszewski, Mikołaj. 1883 [1995]. *Outline of linguistic science*. Amsterdam: John Benjamins Publishing Company.
- Kul, Małgorzata. 2007. "Metaphonology in text messages", in: Barbara Lewandowska-Tomaszczyk – Tomasz Płudowski – Dolores Valencia Tanno (eds), 429-442.
- Kul, Małgorzata – Dafydd Gibbon. 2006. *Channel-driven economy strategies in English and Polish SMS messages*. (Paper presented at the 37th Poznan Linguistics Meetings, 20-23 April 2006, Poznań, Poland).
- Kutschera, Franz von. 1975. "Semantic analyses of normative concepts", *Erkenntnis* 9: 195-218.
- Ladefoged, Peter. 1990. "Some reflections on the IPA", *Journal of Phonetics* 18: 335-346.
- LaVoie, Lisa. 1996. "Consonant strength: Results of a data base development project", *Working papers of the Cornell Phonetics Laboratory*, Volume 11. 269-316.
- LaVoie, Lisa. 2001. *Consonant strength. Phonological patterns and phonetic manifestations*. Outstanding dissertations in Linguistics. New York and London: Garland.
- Lemaître, Georges. 1927. "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extragalactiques" [A homogeneous universe of constant mass and growing radius accounting for the radial velocity of extragalactic nebulae], *Annals of the Scientific Society of Brussels* 47A.
- Lewandowska-Tomaszczyk, Barbara – Tomasz Płudowski – Dolores Valencia Tanno (eds). 2007. *The Media and International Communication. MediaCom 2005*. Peter Lang: Frankfurt an Main. Series: Łódź Studies in Language (edited by Barbara Lewandowska-Tomaszczyk).
- Liberman, Alvin M. 1957. "Some results of research on speech perception", *Journal of the Acoustical Society of America* 29: 117-123.

- Lieberman, Alvin M. – Ignatius G. Mattingly. 1985. "The motor theory of speech perception revised", *Cognition* 21: 1-36.
- Lindblom, Björn. 1990. "Explaining phonetic variation: a sketch of the H&H theory", in: William. J. Hardcastle – Alain Marchal (eds.), 403-439.
- Lindblom, Björn – Ian Maddieson. 1988. "*Phonetic universals in consonant systems*", in: Larry M. Hyman – Charles N. Li (eds.), 62-78.
- Lindblom, Björn – Seung - Jae Moon. 2001. *Energetics of Speech: Proceedings of the International Conference on Speech Processing*, Vol. 2, 841-846. Daejeon, Korea.
- Lisker, Leigh. 1957. "Closure duration and the intervocalic voiced-voiceless distinction in English", *Language*, Vol. 33, No. 1: 42-49.
- Lisker, Leigh. 1972. "Stop duration and voicing in English", in: Albert Valdman (ed), 339-343.
- Löfqvist, Anders – Vincent L. Gracco. 1996. "Labial kinematics in stop consonant production", *The Journal of the Acoustical Society of America*, April 1996, Volume 99, Issue 4, 2472-2500.
- Luce, Duncan R. – Robert R. Bush – Eugene Galanter (eds.). 1965. *Readings in mathematical psychology*, Vol. 2. New York: John Wiley.
- Luschützky, Hans Christian. 1997. *Synchrone und diachrone Aspekte der phonologischen Prozesstypologie [Synchronic and diachronic aspects of process typology]*. Wien: Wien Universitätsverlag.
- Luschützky, Hans Christian. 1999. "Sprachtypologie" [Language Typology], in: Peter Ernst (ed.), 1-54.
- Luschützky, Hans Christian. 2001. "Sixteen possible types of natural phonological processes", in: Charles W. Kreidler (ed.), 506-524.
- Luschützky, Hans Christian. 2005. *Zarys typologii języków [An outline of linguistic typology]*. Kraków: Collegium Columbinum.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: CUP.
- Maddieson, Ian – Kristin Precoda. 1989. "Updating UPSID", *UCLA Working Papers in Phonetics* 74: 104-111.
- Maddieson, Ian. 2005. *Exoticness and complexity: definitions and distributions*. (Paper presented at the 36th Poznan Linguistics Meetings, 22 - 24 April 2005, Poznan, Poland).
- Malmberg, Bertil. 1963. *Phonetics*. New York: Dover Publications, Inc.

- Manaster Ramer, Alexis. 1996. "A letter from an incompletely neutral phonologist", *Journal of Phonetics*, 24: 477-489.
- Mandelbrot, Benoit. 1965. "Information theory and psycholinguistics", in: Benjamin B. Wolman, Benjamin B. – Ernest Nagel (eds.), 550-562.
- Marslen-Wilson, William David. 1987. "Functional parallelism in spoken word-recognition", *Cognition* 25: 71-102.
- Martinet, Andre. 1960. *Elements of general linguistics*. London: Faber and Faber.
- Mateescu, Dan. 2003. *English phonetics and phonological theory: 20th century approach*. (<http://www.unibuc.ro/eBooks/filologie/mateescu/cuprins.htm>) (date of access: 17 March 2005).
- Maupertuis, de Pierre Louis Moreau. 1750. *Essai cosmologie [Essay on cosmology]*. Amsterdam.
- Menzerath, Paul. 1954. "Die Architektonik des deutschen Wortschatzes" [The structure of German vocabularies], *Phonetische Studien*, Heft 3. Bonn: Ferd Dümmlers Verlag.
- Mitton, Roger. 1992. *A description of a computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English*. (<ftp://ota.ox.ac.uk/pub/ota/public/dicts/710/text710.doc>) (date of access: 5 feb 2007).
- Moon, Seung Jae – Björn Lindblom. 1994. "Interaction between duration, context, and speaking style in English stressed vowels", *Journal of the Acoustical Society of America* 96: 40-55.
- Morales, Jose Isidro. 1797. *Memoria matematica sobre el calculo de la opinion en las elecciones [Mathematical memory in assessing the opinion in elections]*. Madrid: Imprenta. (<http://www2.eco.uva.es/~mateco/profesor/lapresta/articulo3.pdf>) (date of access: 27 Nov. 2004).
- Moravcsik, Edith A. 1979. "Typology in Linguistics", *Acta Linguistica Academiae Scientiarum Hungaricae* 29: 315-337.
- Mortimer, Adler J. (ed.). 1991. *The dialogues of Plato*. Chicago: the University Press of Chicago.
- Nadler, Robert D. – James H., Abbs– Osamu Fujimura. 1987. *Speech Movement Research using the New X-Ray Microbeam System: Proceedings: 11th*

- International Congress of Phonetic Sciences*. Tallinn, Estonia, August 1-7, 1987.
- Newton, Isaac. 1687 [1995]. *Newton: texts, backgrounds, commentaries*. ed. I. Bernard Cohen – Richard S. Westfall. New York & London: Norton Critical Edition.
- Ohala, John J. 1984. „Chairman’s summary of Symposium on Phonetic Explanation in Phonology”, in: Marcel P.R. van den Broecke – Antonie. Cohen (eds), 175-182. (http://trill.berkeley.edu/users/ohala/papers/Utrecht_ICPhS_Symposium.pdf) (date of access: 27 Jan 2006)
- Ohala, John J. 1992. “What’s cognitive, what’s not, in sound change”, in Günter Kellerman – Michael D. Morrissey (eds.), 309-355.
- Ohala, John J. 1993. “The phonetics of sound change”, in: Charles Jones (ed), 237-278.
- Oktaba, Wiktor. 1966. „Elementy statystyki matematycznej i metodyka doświadczalnictwa” [The elements of mathematical statistics and research methodology]. Warszawa: PWN [Polish Scientific Publishers].
- Oñederra, Miren Lourdes – Jauregi Ortoiz. 2005. *Fortitive effects of lenition in agglutinative language (Basque)*. (Paper presented at the 38th Societas Linguistica Europaea, 7-10 September 2005, Valencia, Spain).
- Palya, William L. 1985. “Sign-tracking with an interfood clock”, *Journal of the Experimental Analysis of Behavior* 43: 321-330 (<http://www.jsu.edu/depart/psychology/sebac/fac-sch/palya-pevey/palya-pevey.html>) (date of access: 11 April 2005).
- Pareto, Vilfredo. 1897. *Cours d'economie politique [A course in political economy]*. Lausanne/Paris: Rouge.
- Passy, Paul. 1890. *Etude sur le changements phonetiques et leurs caracteres generaux [Studies on phonetic changes and their general characteristics]*. Paris: Didot.
- Peirce, Charles Sanders. 1991. “The collected papers”, in: James Hoopes (ed.).
- Petrova, Olga. 2001. *The evolution of English obstruent system: an optimality theoretic approach*. [PhD dissertation, Iowa University]. (<http://www.uiowa.edu/~linguist/petrova.html>) (date of access: 6 May 2004).
- Pierrehumbert, Janet – David Talkin. 1992. “Lenition of /h/ and glottal stop”, in: Gerard J. Docherty – D. Robert Ladd (eds.), 90-117.
- Pierrehumbert, Janet – Michael Broe (eds.). 2000. *Papers in laboratory phonology*, vol.6. Cambridge: CUP.

- Plato. 360 BC [1991]. "Cratylus", in: Adler J. Mortimer (ed.), 89-99.
- Poupplier, Marianne. 2003. "The dynamics of error", in: Maria Josep Solé – Daniel Recasens – Joaquín Romero (eds.), 2245-2248.
(<http://www.speech.umaryland.edu/~marianne>) (date of access: 13 Feb. 2005).
- Reichle, Eriks D – Patricia A. Carpenter – Marcel A. Just. 2000. "The neural basis of strategy and skill in sentence-picture verification", *Cognitive Psychology* 40: 261-295. (<http://www.ccbi.cmu.edu/reprints/reprints.htm>) (date of access: 10 April 2005).
- Rescher, Nikolas. (ed.). 1986. *Current issues in teleology*. Lanham MD: University Press of America.
- Ronneberger-Sibold, Elke. 1984. "Nhd. quer, Zwerchfell und <<natürlicher Wandeln>>" [Low - high German (Niederhoch Deutsch) across diaphragm and <<natural changes>>], *Wiener Linguistische Gazette*, Supplement 3: 221-225.
- Saltzman, Elliot L. – Scott, J.A. Kelso. 1987. "Skilled actions: A task dynamic approach", *Psychological Review* 94: 84-106.
- Sapir, Edward. 1921 [1949]. *Language. An introduction to the study of speech*. New York: Harcourt, Brace & World, Inc.
- Saussure, Ferdinand, de. 1915 [1983]. *Course in General Linguistics*. Translated by W. Baskin. London: Duckworth.
- Schiffrin, Deborah. 1984. *Meaning, form, and use in context: linguistic applications*. Washington, DC: Georgetown University Press.
- Selkirk, Elizabeth O. 1980. "The role of prosodic categories in English word stress", *Linguistic Inquiry* 11 (3): 563-605.
- Sen, Amartya Kumar. 1964. "Preferences, votes and the transitivity of majority decisions", *Review of Economic Studies* 31: 163-165.
- Sharma, Ram Nath. 1987. *The Astadhyayi of Panini*. Vol I. New Delhi: Munshiram Manoharlal Publishers Pvt. Ltd.
- Sharma, Ram Nath. 1990. *The Astadhyayi of Panini*. Vol II. New Delhi: Munshiram Manoharlal Publishers Pvt. Ltd.
- Shockey, Linda. 2003. *Sound patterns of spoken English*. Oxford: Blackwell Publishing.
- Sievers, Eduard. 1876. *Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der Indogermanischen Sprachen [The fundamentals of sound physiology in the introduction to study of sounds theory in the Indogermanic language]*. Leipzig: Breitkopf und Härtel.

- Sievers, Eduard. 1893 [1901]. *Grundzüge der phonetik [The fundamentals of phonetics]*. Leipzig: Breitkopf and Härtel.
- SIL International. 1999. *What is the phonological hierarchy?*
(<http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsThePhonologicalHierarchy.htm>) (date of access: 10 April 2005).
- SIL International. 2004. *What is the sonority scale?*
(<http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsTheSonorityScale.htm>) (date of access: 27 February 2007).
- Silverman, Daniel. 1997. *Phrasing and recoverability. Outstanding dissertations in linguistics series*. New York: Garland Publishing.
- Silverstein, Michael (ed.). 1971. *Whitney on language. Selected writings of William Dwight Whitney*. Cambridge, Massachusetts and London, England: The MIT Press.
- Simon, Fritz B. – Helm Stierlin. 1998. *Słownik terapii rodzin [The language of family therapy. A dictionary of terms]*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne, 97-100, 423-427.
- Simpson, Rita C. – Sarah L. Briggs – Janine Ovens – John M. Swales. 2002. MICASE. The Michigan Corpus of Academic Spoken English. Ann Arbor, MI: The Regents of the University of Michigan.
<http://www.lsa.umich.edu/eli/micase/index.htm>
- Slobin, Dan Isaac. 1985a. *The crosslinguistic study of language acquisition: Vol. 1. The data*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan Isaac. 1985b. *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Słownik Wyrazów Obcych [A dictionary of foreign words and phrases]*. 1997. Warszawa: Wydawnictwo Naukowe PWN [Polish Scientific Publishers].
- Solé, Maria Josep – Daniel Recasens – Joaquin Romero (eds.). 2003. *Proceedings of the 15th International Congress of the Phonetic Sciences*. Rundle Mall: Causal Productions.
- Sommerstein, Allan H. 1977. *Modern phonology*. London: Edward Arnold.
- Sotillo, Cathy. 1997. *Phonological reduction and intelligibility in task-oriented dialogue*. [PhD dissertation, University of Edinburgh].
(<http://www.cogsci.ed.ac.uk/~cathy/>) (date of access: 17 Feb. 2004).

- De Spinoza, Benedictus. 1662, 1667 [1955]. *On the Improvement of the Understanding, The Ethics, Correspondence*. (Translated by R. H. M. Elwes). New York: Dover Publications.
- Stampe, David. 1969 [1979]. *The acquisition of phonetic representation*. New York and London: Garland Publishing, Inc.
- Stampe, David. 1973 [1979]. *A Dissertation on Natural Phonology*. New York and London: Garland Publishing, Inc.
- Stampe, David. 1973. "On Chapter Nine", in: Michael Kenstowicz – Charles W. Kisseberth (eds.), 44-52.
- Steriade, Donca. 2000. "Paradigm uniformity and the phonetic/phonology boundary", in: Janet Pierrehumert – Michael Broe (eds.), 313-335.
- Sweet, Henry. 1891 [1960]. *A new English grammar, logical and historical*. Part I: Introduction, Phonology, and Accidence. Oxford: At the Clarendon Press.
- Szaniawski, Klemens. 1994. *O nauce, rozumowaniu i wartościach. Pisma wybrane [On science, thinking and values. Selected writings]*. Warszawa: Wydawnictwo Naukowe PWN [Polish Scientific Publishers].
- Trask, Robert Lawrence. 1993. *A dictionary of grammatical terms in linguistics*. London & New York: Routledge.
- Trask, Robert Lawrence. 1996. *A dictionary of phonetics and phonology*. London & New York: Routledge.
- Trnka, Bohumil. 1936 [1982]. "General laws of phonemic combinations", in Fried Vilem (ed.), *In Selected papers in structural linguistics. Contributions to English and general linguistics written in the years 1928-1978*. ed. by Vilem Fried. Berlin: Mouton Publishers. 113-119
- Trubetzkoy, Nikolai. 1931. "Die phonologischen Systemen" [The phonological systems]. *TCLP* 4: 96-116.
- Tsai, L.S. [full name not found]. 1932. "The laws of minimum effort and maximum satisfaction in animal behavior", *Monograph of the National Institute of Psychology*, no 1.
- Valdman, Albert (ed). 1972. *Papers in linguistics and phonetics to the memory of Pierre Delattre*. The Hague, Paris: Mouton.
- Vennemann, Theo. 1983. "Causality in language change. Theories of linguistic preferences as a basis for linguistic explanations", *Folia Linguistica Historica* VI/1: 5-26.

- Vennemann, Theo. 1988. *Preference laws for syllable structures*. Berlin: Mouton de Gruyter.
- Vihman, Marilyn. 1977. *A Reference manual and user's guide for the Stanford Phonology Archive*. Part I. Stanford University.
- Volk, William. 1965. *Applied statistics for engineers*. Warszawa: WNT.
- Waniek-Klimczak, Ewa – Włodzimierz Sobkowiak (eds.). 2003. *Dydaktyka fonetyki języka obcego*. Neofilologia, tom V. Płock: Zeszyty Naukowe PWSZ w Płocku.
- Waters, R.H. [full name not found].1937. "The principle of least effort in learning", *Journal of General Psychology*, Vol. 16. 3-20.
- Watson, Kevin. 2006. *The phonetics and phonology of plosive lenition in Liverpool English*. [PhD dissertation, Lancaster University].
- Wheeler, R. H. [full name not found].1929. *The science of psychology*. New York: Crowell.
- Whitney, William Dwight. 1878 [1971]. "The principle of economy as a phonetic force", in: Michael Silverstein (ed.), 249-260.
- Wolman, Benjamin B. – Ernest Nagel (eds.). 1965. *Scientific Psychology: principles and approaches*. New York: Basic Books.
- Yule, George Udny. 1944. *Statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
- Zipf, George Kingsley. 1935 [1965]. *The psycho-biology of language*. Massachusettes: the MIT Press.
- Zipf, George Kingsley. 1949 [1972]. *Human behavior and the principle of least effort. An Introduction to human ecology*. New York: Hafner Publishing Company.