

KATARZYNA DZIUBALSKA-KOŁACZYK
WILLIAM A. KRETZSCHMAR, JR.

COMPLEX NATURAL SYSTEMS FOR LANGUAGE

INTRODUCTION

In this paper we claim that complexity science (which deals with complex systems) can enhance the theory of natural phonology. This leads us to propose complex natural systems modelling for language. In the analysis, we bring together the perspectives of sociolinguistics and language acquisition as well as phonotactics. Empirical observations from those areas constitute a bottom-up motivation for complex systems analysis. On the other hand, extralinguistic principles for language, such as cognitive or semiotic ones, which are fundamental for natural phonology, are theoretical inspirations for research on linguistic complexity. In complex systems structures emerge without any specific cause (Ellis & Larsen-Freeman, 2009; Kretzschmar, 2009), while in natural linguistics they result from functional behavior and teleology of change (Dressler & Dziubalska-Kołaczyk, 1994). In complex systems frequency of usage drives a structure towards stability (Mitchell, 2009; Kretzschmar, 2015; Burkette, 2016) while in natural linguistics a degree of preferability is a major force behind survival of structures. Universals are understood as post-factum generalizations in complex systems (where the only true universal is human interaction) while they derive from general semiotic and functional principles in natural linguistics. Sociological and psychological factors are interpretable and matter in both complex systems and natural theory. Despite the apparent divergences,

Prof. Dr Hab. KATARZYNA DZIUBALSKA-KOŁACZYK, Full Professor at Adam Mickiewicz University, Poznań, Poland; e-mail: dkasia@amu.edu.pl; ORCID: <https://orcid.org/0000-0003-4884-3448>.

WILLIAM A. KRETZSCHMAR, JR., PhD, Willson Professor in Humanities at the University of Georgia, USA; e-mail: kretzsch@uga.edu; ORCID: <https://orcid.org/0000-0002-7173-5624>.

we will demonstrate the applicability of the complex systems apparatus to the natural linguistics framework (cf. Dressler, 2011).

1. COMPLEX SYSTEMS

When new ideas in linguistics become available, those who follow existing models need to address them. One way to do so is to reject anything new, a method often used by those with entrenched positions but one which does not allow any more than incremental improvements in the field. Another way to do so is to throw over the older model in favor of the new one, as when generativism swept linguistics in the last century, but just abandoning older ideas loses all the useful thought and development of our predecessors. In this paper, we hope to take a middle way by considering how the new idea of complex systems in linguistics may apply to the existing model of natural linguistics. We will not just mix the two, like mixing black and white paints to make gray, which would lose the distinctive advantages of both the newer and the older model. Instead, we will argue that there is a specific relationship between complex systems and natural linguistics which improves our understanding of each model. We explain how complex systems help us understand both theoretical concepts from natural phonology, for example the concept of preference, as well as variation in language data, for example in vowel and consonant cluster realizations.

Let us begin with complexity science. In Mitchell's (2009, p. 13) definition, a complex system is "a system in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution." The new science of complex systems (CS), also known as complex adaptive systems or complex physical systems (or complex dynamic systems in applied linguistics), was launched in 1984, when the Santa Fe Institute was founded for its study. CS were originally described and are still used in the physical and biological sciences (e.g. Prigogine & Stengers, 1984; Hawking & Mlodinow, 2010; Gould, 2003), somewhat later in computer science (e.g. Holland, 1998). CS received early allusive discussion in linguistics: Lindblom, MacNeilage, and Studdert-Kennedy published in 1984 a paper on self-organizing processes in phonology; Paul Hopper presented his seminal paper called "Emergent Grammar" in Berkeley in 1987; Ronald Langacker published a chapter on "A Usage-Based Model" for cognitive linguistics in 1988. Larsen-Freeman (1997) suggested complexity science for the study of language acquisition. Ellis and Larsen-Freeman (2009) discovered the nonlinear pattern for ESL in *Language as a Complex Adaptive System* (for the most recent review of research

methods for complexity theory in applied linguistics and Complex Dynamic Systems Theory see Hiver & Al-Hoorie, 2020). Work by Joan Bybee (2001, 2002) promoted the importance of frequency of use and eventually referred to CS (2010). In Bybee's interpretation of complex systems, substance (i.e., phonetics and semantics) and use interact to create structure. The pillars of Bybee's well-established usage-based approach to language are frequency effects, creativity of repetition, and the notion of schemas or emergent generalizations. Three recent books, however, have embraced CS and developed ideas about it much more fully. Kretzschmar (2009) has demonstrated how complex systems do provide the underlying pattern for speech in *The Linguistics of Speech*, focusing on nonlinear distributions and scaling properties. Kretzschmar (2009, Chapter 6) connects current ideas about complex systems with Mandelbrot's earlier work on fractals. Mandelbrot had worked previously on Zipf's Law and modified it in light of detailed study of evidence (1968, 1982), and so created the association of language with fractal distributions. Kretzschmar (2009) located Mandelbrot's idea of fractals within the newer field of CS (which offers a number of possible mathematical models, of which Mandelbrot's fractals are the best fit for language), and established that a wide range of linguistic data follows distributional predictions based on nonlinear distributions with self-similar scaling. Kretzschmar's (2015) *Language and Complex Systems* applies CS to a number of subfields in linguistics. Finally, Burkette's (2016) *Language and Material Culture: Complex Systems in Human Behavior* applies CS to both the study of language and the anthropological study of materiality. There is also now an undergraduate textbook, *Exploring Linguistic Science* (Burkette & Kretzschmar, 2018), that offers an easier pathway to introduce CS to linguists including chapters especially for corpus linguists.

Let us think now about language as a complex system. This is not how we have all been taught to think about language. All through the earlier years of school, our teachers always told us about the grammar of our language, rules that we should follow and violate only at our peril. Sometimes our teachers told us about dictionaries, where the dictionary was the authority that gave us the legal words in the language along with their meanings. Grammatical rules and lists of words in the dictionary provided structure for our language, syntagmatic and paradigmatic systems, respectively. It is no surprise, then, that we think of language as a structured entity, where we can decide what is grammatical or not according to the syntagmatic structure of the language, and where we know where to find the list of legal words, the paradigmatic structure of the language. These ideas also inform most modern approaches to linguistics.

Syntagmatic and paradigmatic structures are not the place to start when we think about language as a complex system. Instead, let us consider that people just use

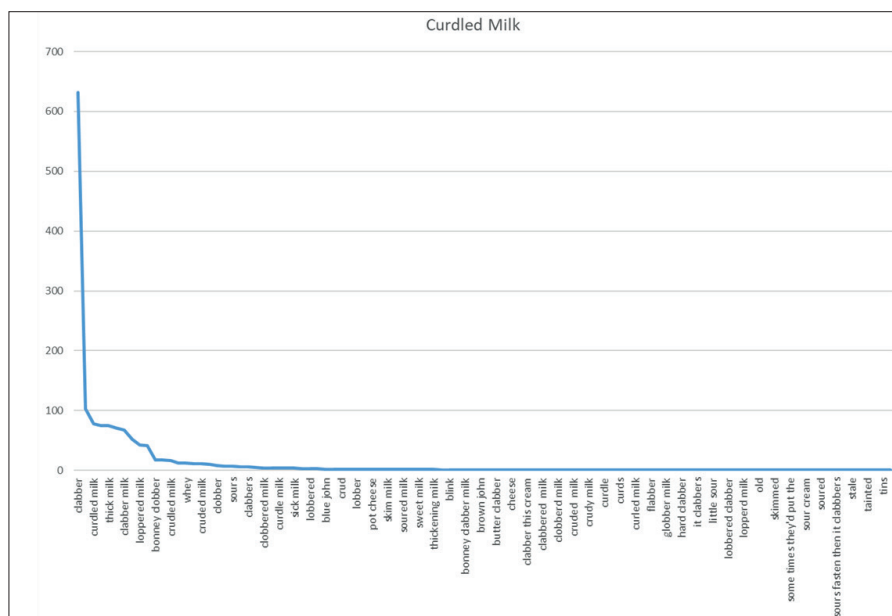
whatever experience they have in order to talk with each other. People are talkers. A person is not the same thing as a language: some people can use their experience to talk in two or more languages. Indeed, all of us do that even if we do not realize it: we all use somewhat different language in all of the different conversations that we have. Our conversations with family members have different words and different grammar than the conversations we have at work. Conversations we have with children have different words and grammar than conversations we have with doctors, lawyers, and other professionals, and we all know what to do to talk with children if not always with professionals. Those of us who write articles and books prepare them for different audiences: we should not write the same article for an audience of engineers or computer scientists that we would write for linguists. The words and grammar are not completely different in these various conversations and situations for writing, just more or less likely that the talker or writer would use particular features in a particular situation. We probably would not talk about “syntagmatic and paradigmatic structure” with members of our families, but we have just done so for this audience because we expect this audience to be comfortable with those words. Each of us speaks a somewhat different language in all of the settings for speech and writing that we encounter, based on our experiences with such settings.

All of that happens because of the complex system of language. We speakers and writers are *agents*, users of language, in the same way that buyers and sellers are agents, users of money. The *components* in the complex system of a language are all possible variant realizations of linguistic features, different word and grammar choices we use for children and professionals, for engineers and linguists, and also different pronunciations. The *activity* in the system consists of all our conversations, and as a more limited case, all our writing. The *exchange of information* is not the same as sharing the meaningful content of what we say and write (which is exchange in a different sense), but instead the implicit comparison of the use of different linguistic variants by different agents in different situations. *Feedback* from exchange of information causes *reinforcement*, so speakers and writers are more likely to use particular variants in future occurrences of the same particular circumstances for speech. Feedback and reinforcement create the nonlinear patterns that appear in linguistic data, in which a few variants are very common, some variants are moderately common, and most variants are rare (when graphed, the long-tailed or fat-tailed pattern described below as an asymptotic hyperbolic curve) at every level of scale. That is, human speakers learn how to talk in every situation of use, a little differently in different situations. Human agents can think about and choose what linguistic variants to use, but that does not change the fact that we make choices in relation to the system, usually without thinking too much about them. The order that emerges in speech is simply the configuration of components, whether words, pronunciations,

or grammatical constructions, that come to occur in all of the circumstances in which we actually communicate. As for other complex systems, they are all conditioned by contingency and so *emergent patterns in language* (languages like English or Polish at the top level of scale and different varieties at a lower level of scale) can change over time. The process operating in complex systems just explains better what we already knew: we tend to talk like people nearby, either physically or socially near, and we tend to use the same linguistic tools that others do when we are writing or saying the same kind of thing.

The operation of the complex system leaves behind a characteristic distributional pattern. Every linguistic feature, whether words or pronunciations or grammatical constructions, has many more variants than one might expect. The frequency profile of all the variants for any feature always forms the same sort of graph, an asymptotic hyperbolic curve (or A-curve for short). Figure 1 shows this pattern for the 105 variant terms for ‘curdled milk’ collected in a survey of 1162 speakers from the American Eastern States, of which the ten most frequent forms are shown.

Figure 1
Terms for ‘Curdled Milk’ in the American Eastern States

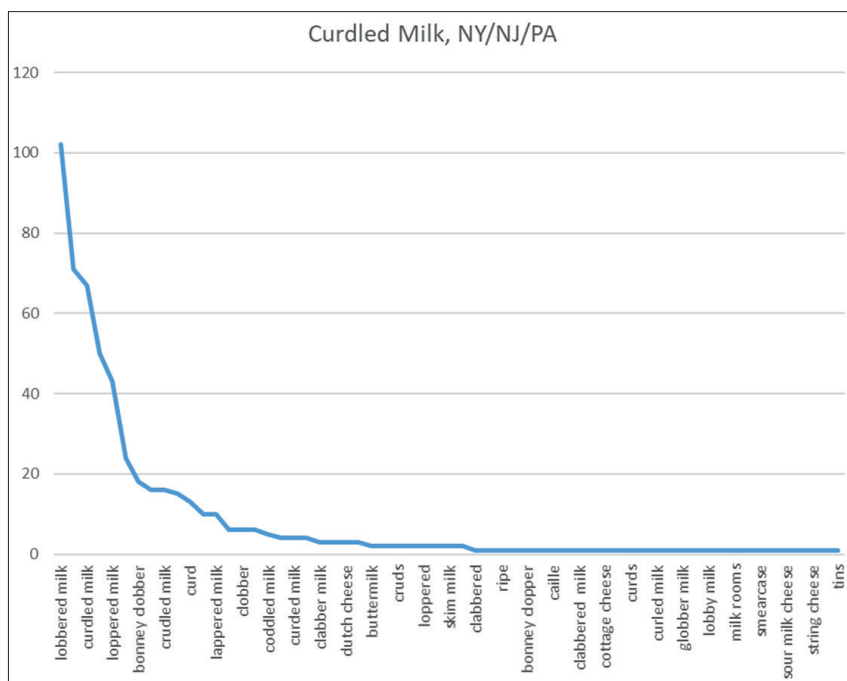


clabber	632	lobbered milk	103
curdled milk	78	clabber milk	68
sour milk	75	loppered milk	43
thick milk	75	clabbered milk	41
bonney clabber	71	bonney clobber	18

A few of the variants, like *clabber* or *lobbered milk*, are very common in the peak of the curve at the left, and a few variants like *bonney clabber* in the middle of the curve occur sometimes, but most of the variants in the long tail of the curve at right were only elicited once. Some people will want to associate this A-curve with Zipf's Law (Zipf, 1949), the idea that rank and frequency are inversely proportional for words in a text, but the pattern is pervasive throughout every aspect of language observed so far, not just for words in texts. Kretzschmar (2009) shows that the A-curve pattern occurs not just in vocabulary variation but also in pronunciation variation and among collocations in corpora and cites examples of the occurrence of the pattern from a great many world languages. The A-curve distribution is an example of what is generally known as a "power law" in fields outside of language study, which has generated substantial comment; the association of the A-curve with CS for language data helps to insulate the distributional pattern from the wide range of speculations found in the power law literature. The nonlinear shape of the frequency profile is similar but not exactly the same in different experiments, so the pattern is not the result of a formula as Zipf proposed, not a law per se as in physics (see Kretzschmar, 2009, Chapter 6 for statistical evaluation of such distributions, and Kretzschmar, 2015, Chapter 7 for curve-fitting in comparison to normal distributions). The term "nonlinear" here corresponds to the description of Mandelbrot (1968, 1982), for which the raw counts form an A-curve and a chart that is logarithmic on both axes yield a straight line. However, the nonlinear A-curve, as differentiated from the bell curve of normal distributions, regularly emerges from the linguistic interactions of speakers in the complex system. As described in detail in Kretzschmar (2015, pp. 182–184), A-curve distributions are essentially different from Gaussian normal distributions, as shown by the Gini Coefficient. Normal distributions plotted by frequency rank may appear to be curvilinear, especially as standard deviation increases (extremely high standard deviations may make a plot appear nonlinear in the sense described here), but the Gini Coefficient shows that they are not nonlinear in the same sense as distributions showing the A-curve pattern. A-curves are not just left-skewed Gaussian distributions.

Moreover, the same A-curve pattern occurs for whatever subsection of the data one might want to consider (the self-similar scaling property is also described in detail in Mandelbrot, 1982). Figure 2 shows the top ten list and the A-curve in the states of New York, New Jersey, and Pennsylvania.

Figure 2
Terms for ‘Curdled Milk’ in the States of New York, New Jersey, and Pennsylvania



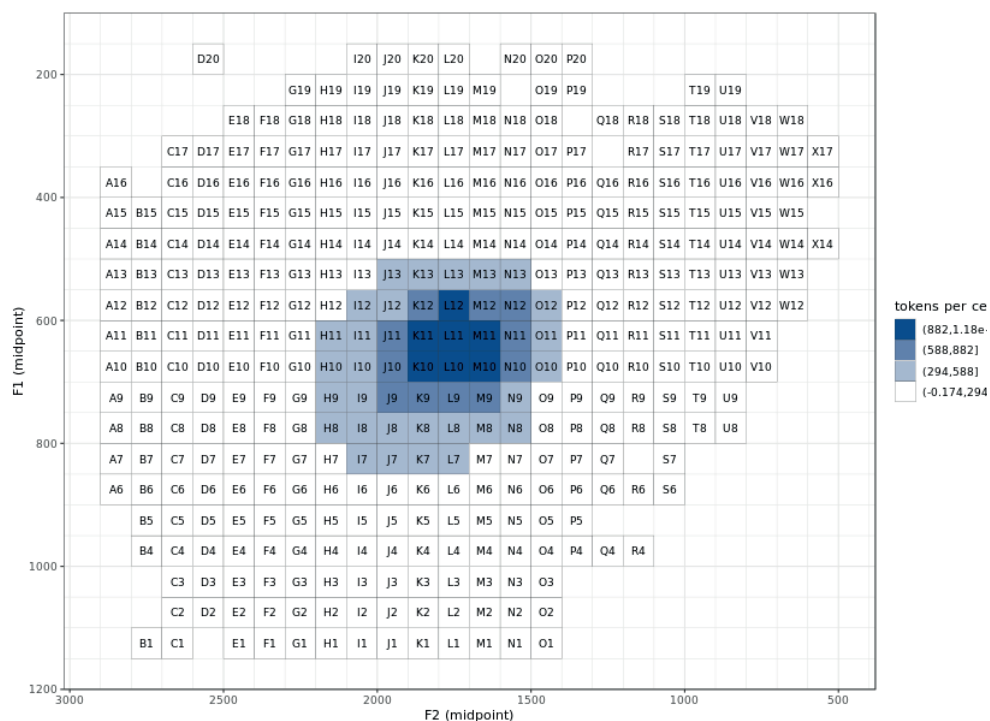
lobbered milk	102	bonney clabber	24
thick milk	71	bonney clobber	18
curdled milk	67	clabber	16
sour milk	50	curdled milk	16
loppered milk	43	clabbered milk	15

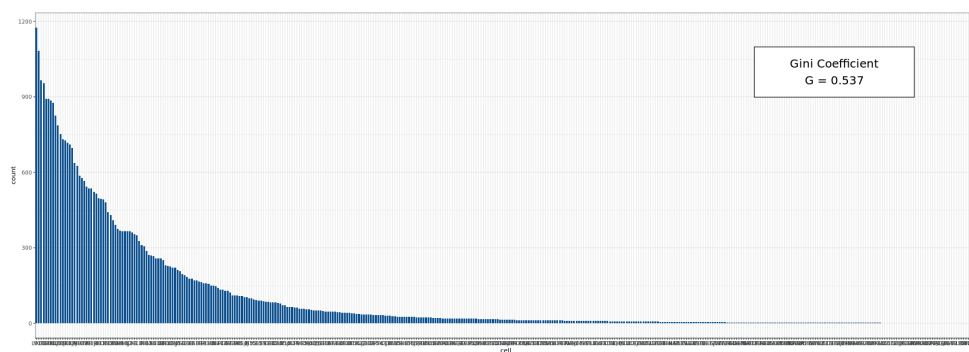
The graph only has 61 variants, not 105, but the peak at the left and the long tail at right are clearly visible. What is different about the overall A-curve and the New York/New Jersey/Pennsylvania A-curve is the order of the variants on the curve; for example, *clabber* is only in 8th place in New York/New Jersey/Pennsylvania, while it was top-ranked in the survey overall. This difference in the frequency of individual variants on the curve, whether between *clabber* and *lobbered milk* to distinguish regional speech or between words and grammatical constructions we use with children and professionals of different kinds, is how we can tell apart the speech of different areas or different social groups or different professional groups. The difference in frequency is how we talk differently with family members and children and different audiences. These groups of talkers are no different from regional and social and professional groups: every group has its own A-curves for words and pronunciation and grammatical choices, with mostly the same variants in a different

order along the curve. The scaling property of the A-curve (confusingly, also known as the scale-free property, in the same way that *inflammable* can mean ‘not able to catch fire’ at the same time it means ‘able to catch fire’) allows us to recognize an unlimited number of different groups according to the frequency rankings of the A-curve, and we talkers use this property all the time to employ the most appropriate language for any situation.

Finally, we can illustrate that pronunciation features have the same distributional properties as words. Figure 3 shows an F1/F2 chart for the realizations of the /æ/ vowel by 63 Southern American speakers in an automatic vowel measurement experiment (funded by the American National Science Foundation, see Olsen et al., 2017; anyone can make plots like these with an online tool created by Joseph Stanley, at lap3.libs.uga.edu/u/jstanley/vowelcharts/). There are 40304 tokens plotted, all under primary stress with no normalization or filtering.

Figure 3
/æ/ Realizations From 63 Southern American Speakers



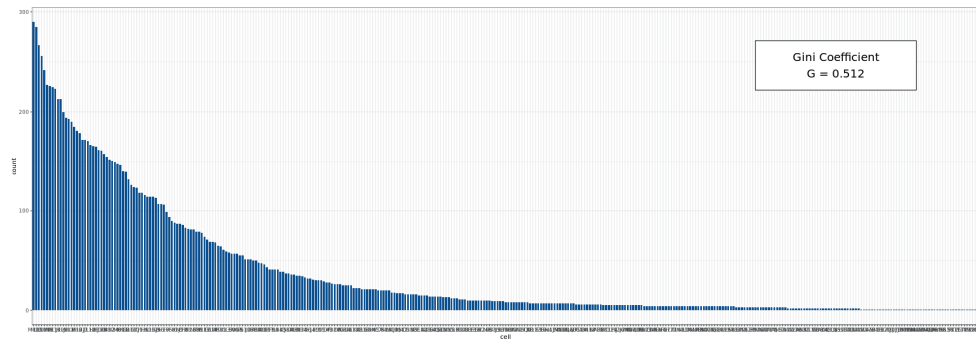
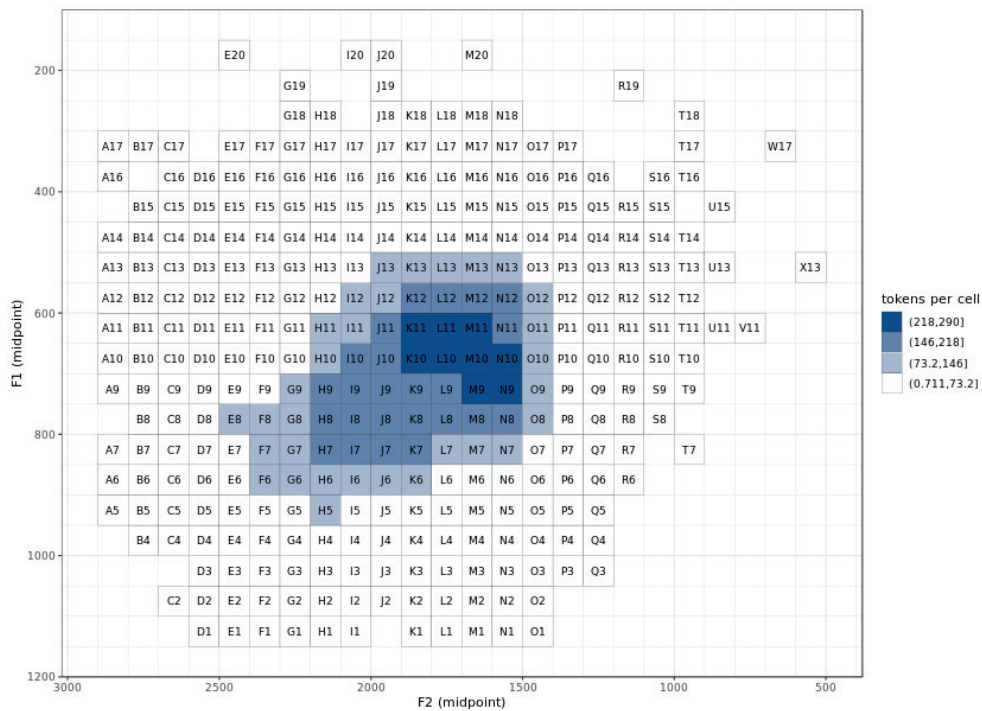


The F1/F2 plot has been arranged in a 20 x 24 grid (480 cells), and the number of tokens located in each cell has been tabulated. Cells with no tokens have no labels (there are some realizations across most of the possible cells); cells with tokens are colored in quartiles, with the most densely populated cells in dark shading and other cells in lighter shading. The mean value for the plot occurs in cell K10, the bottom leftmost of the dark cells, so the distribution is not accurately described by the mean. Rather than the central tendency (centroid) expected in a normal distribution, the densities of cells occur in a nonlinear A-curve: the chart of frequencies per cell below the plot shows the A-curve, and the Gini Coefficient measures the sharpness of the curve, less sharp than Zipf's Law would predict but distinctly nonlinear (Kretzschmar, 2015, Chapter 7 shows that the Gini Coefficient for a normal distribution is typically very low, at 0.1 or lower). Figure 3 shows that while /æ/ can be realized widely across F1/F2 space for these speakers there is a range of cells for what people usually say (in dark shading), and some other cells in lighter shading for what people often say.

Figure 4 shows the F1/F2 plot and A-curve for 13318 tokens just from the most highly educated subgroup of speakers. As was the case with the subgroup of speakers for variants of 'curdled milk,' the subgroup of educated speakers shows a similar nonlinear frequency profile, and also as for 'curdled milk', the profile has a somewhat different order of cells. Now the densest (darkest) pattern is somewhat lower and more back, and the mean value for F1/F2 occurs in cell K9, outside of the darkest cells, so the central tendency is again not a good representation of the heart of the distribution. Pronunciation evidence, then, follows evidence from words in showing the characteristic distributional pattern created by the complex system, including its scale-free property that shows a similar distributional pattern at any level of scale.

Figure 4

/æ/ Realizations From 18 Southern American Highly Educated Speakers



It is of course possible to be heavily involved with mathematics when we study complex systems. The usual kind of advanced mathematics taught in modern schools, calculus, works well with motion as when Isaac Newton invented it to describe the motion of the planets. Calculus does not, however, work with everything, and it does not describe the nonlinear and scaling effects of complex systems. Fractal mathematics, as named by the mathematician Mandelbrot (e.g. 1982), is better for many aspects of the natural world, such as coastlines, the branching structure of trees, and language evidence. The A-curve and the scaling properties of language are the best signs that language is fractal, as it arises from a complex system of interactions.

Even without difficult mathematics, the distributional patterns left by the complex system of speech are clearly visible when plotted, whether in F1/F2 charts or as A-curves. These patterns all derive directly from evidence, from counting tokens of the possible variants for some recognizable feature of language. They are thus very concrete. However, we should not forget that the complex system of speech depends essentially on interactions between speakers, which creates feedback and reinforcement that eventually yields these emergent patterns. Constant activity in the complex system is required for its operation; without it we have only dead languages, like Latin, which cannot maintain emergent patterns without a large number of speakers to interact with each other. Nonlinear frequency profiles, therefore, are an important part of the story but not the whole story, and we must also have scope to describe the operation of interactions.

2. NATURAL LINGUISTICS

Let us now turn to natural linguistics. Natural linguistics began as natural phonology in the works of David Stampe (1969, 1979) and then joint works of David Stampe and Patricia Donegan (1979 and later). Natural phonology (NP) was proposed as an alternative to both structural and generative approaches to phonology current at the time. It differed from the other approaches by a fundamental idea that phonological systems were phonetically motivated. According to Stampe, a child is born with an infinite potential for universal phonological processes which, in the course of acquisition, undergo suppression, reordering or limitation informed by the ambient speech. So, for example the process of syllable-final obstruent devoicing is fully suppressed in English, it is limited to word-final position in Polish, and it applies fully in standard German. In other words, a child overcomes phonetic difficulties of production and perception by means of phonetically motivated processes of lenition and fortition. This view of acquisition allows for both convergencies and divergencies in the speech of children acquiring a given language as well as across languages,

since, while there is one phonetics, languages choose from it differently. Lenitions and fortitions relate directly to ease of articulation (speaker-friendly context-sensitive processes) and clarity of perception (listener-friendly context-free processes), later referred to as foregrounding and backgrounding processes, respectively, by Dressler (1985, 1996). Lenitions are especially well-attested in casual (normal everyday) style in which ease of articulation benefits the speaker. Although the original assumption of NP was that natural, universal processes were innate, Donegan (1985) noted that “it would not alter the theory of natural phonology substantially to say that processes may be discovered by the child as he learns to use his vocal tract” (p. 26, note 5). In other words, processes may be understood as emergent in response to the difficulties posed to a child trying to make efficient use of the inborn capacity for articulation and perception of linguistic sounds. Processes emerge universally; this, however, does not imply that they are identical for all children. Since children are active in acquisition, and they are influenced by a particular ambient language, they discover divergent solutions to the difficulties they face, retreat from already entered paths, and so on.

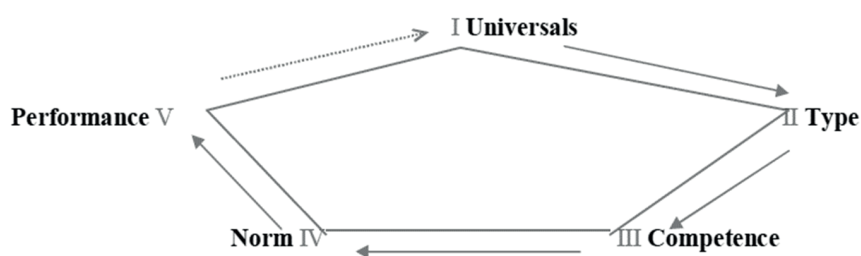
Natural phonology was subsequently expanded to become a full-fledged theory of language known as natural linguistics (NL). This development has been largely due to the extensive research in natural morphology and morphonology, sociolinguistics, pragmatics, text linguistics and historical linguistics by Wolfgang U. Dressler (1985, 1996 and hundreds of other publications) as well as his disciples (for overviews see Dziubalska-Kołaczyk, 2001, 2002, 2006, 2008, 2012). Natural linguistics is a cognitive functional theory of language with semiotic metatheoretical underpinnings. It is a preference theory in the sense that generalizations do not acquire the status of absolute laws but form hierarchic continua from the most preferred through dispreferred, explicable in terms of prototypes and defaults. A human agent—a language user—is the lens of a linguistic system. The external circumstances of the speaking situation are as important for the shape of linguistic output as the internal grammatical evidence. As Stampe implied, competence is the competence of performance (Stampe, 1969, 1979). Therefore, studies in natural linguistics take into account external evidence from language change, variation, acquisition and use. As a consequence, corpora of performance data constitute verification ground for natural linguistic ideas. Speakers differ from other speakers and intra-individually in the way they use language depending on multiple factors: age, state (tired, drunk, sick, aphasic), geographical area, social status, gender, personal traits, social context of communication and so on.

Naturalist theory is predictive, although its explanations are not exactly deductive-nomological, since the latter are more suitable “for (relatively) closed systems” (Dressler, 1985, p. 289), such as in physical sciences. Still, controlled experiments

are possible also in linguistics. The actual epistemological framework of NL is functionalist, with the reservation of plurifunctionality. For example, the process of vowel epenthesis in second language acquisition of English by Japanese learners serves both production and perception (sequences without clusters are easier to pronounce and they also facilitate perception). However, consonant deletion would bring the same effect. Functional goals may also conflict, and thus a hierarchy of functions is implied as well as the observation that form follows function only to some extent. Thus, processes (expressed as phonological, morphological, syntactic rules) on all levels of the explanatory model proposed by Dressler (1985, pp. 292ff) serve functions (such as ease or clarity). The explanatory cycle of natural linguistics is very adequately visualized by Dressler's (1985) quintuple.

Figure 5

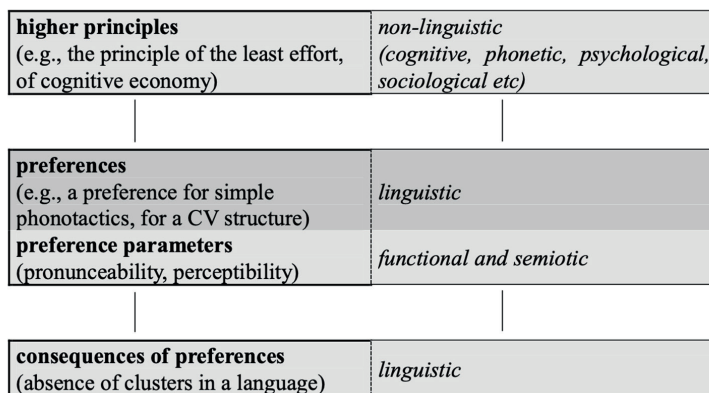
The Universals-to-Performance Quintuple



Note. Reproduced from Dressler, 1985, p. 292.

The quintuple (originally established by Hjelmslev and Coseriu, cf. Dressler, 1985, p. 292) has been adapted by Dressler to replace the Chomskyan triple (I, III and V) and the Saussurean quadruple (I, III, IV and V), since it shows the path from universal properties of language to individual performance in steps compatible with the naturalist framework. Each of the five elements is simultaneously the basis of and is filtered by the next one. One needs to consider the whole quintuple when attempting to account for performance. The levels in the diagram represent three subtheories of naturalness, i.e. (I) the theory of universals (of the human language faculty), (II) the theory of type adequacy and (III) the theory of language-specific system adequacy. (IV) normative, i.e. sociolinguistic factors and (V) psycholinguistic factors further contribute to the final shape of performance. Performance, in turn, has the potential to modify universals.

From a still wider perspective, explanation in NL takes the path starting with higher, non-linguistic principles, via linguistic preferences through their consequences in specific languages (cf. Figure 6 below).

Figure 6*Explanatory Schema in NL*

Dressler adopted Peircean semiotics (Peirce, 1965), which is functional and processual, as a metatheory of NL. Semiotically based preferences are those for iconicity, indexicality, (bi)uniqueness, figure & ground sharpening, binarity. In the above schema, semiotic preference parameters are mentioned, since semiotics deals in a gradual way with the parameter: natural β à conventional and thus allows for gradual parametrization. What language and semiotics have in common are signs: a theory of signs seems to be a good choice to describe the system of signs.

Let us demonstrate the way in which a linguistic phenomenon is explained using the tools of natural linguistics. Our example will concern the behaviour of consonant clusters, described by phonotactics and morphonotactics. Phonotactic grammar is concerned with well-formedness of consonant clusters and operates on basic, non-derived, lexical forms (e.g., the final clusters in *band* and *past*). Morphonotactics takes care of the remaining, morphologically complex, forms (e.g., the final clusters in *ban(n)+ed* and *pass+ed*). Morphonotactics is the area of interaction between morphotactics and phonotactics (Dressler & Dziubalska-Kołaczyk, 2006) and shows how inflection, word-formation and compounding contribute to the creation of consonant clusters.

How do we go about predicting which clusters are preferred by speakers? Such predictions should be universally relevant across languages and across domains, i.e. they should work for language lexicon, language use, acquisition and change. We start with a general principle of perception: contrast. The Gestalt-psychology principle of figure and ground (first investigated by the Danish psychologist Edgar Rubin, cf. Rubin Face/Vase Illusion, named after him) captures the idea. Thus, “to construct a useful signalling system out of sound, there must be some differentiation between different parts of the signal in time” (Maddieson, 1999, p. 2525). As

is well established, the sequences whose elements are differentiated best from each other are consonants followed by vowels (CV). Typological evidence shows that CV sequences are the only ones that occur in all languages of the world. Children acquire CV sequences before more complex ones, and normal speech is ridden with reduction processes towards CV. Also historically, CV sequences are more stable than others, because “larger modulations have more survival value than lesser ones and therefore will persist in languages” (Ohala, 1990, p. 326).

At the same time, however, languages that have only CV sequences constitute a minority of only about 12.5% (Maddieson, 2009), 56.6% allow moderately complex CCVC sequences as well, and 30.9% allow more complex ones, some of them even sequences as complex as CCCVCCCC. Thus, it is obvious that there are conditions under which consonant clusters can come to be established in natural languages, even though they may be dispreferred on perceptual grounds.

The most straightforward prediction is that the more complex (the longer) a cluster is, the less preferred it appears. Next, we propose that a cluster’s preferability reflects the strength of the contrasts between cluster constituents (cf. NAD—Net Auditory Principle,¹ Dziubalska-Kołaczyk, 2014, 2019). We also predict preferred clusters to be more frequent than dispreferred ones. In relation to morphotactics, we expect relatively marked (dispreferred) clusters across morpheme boundaries and relatively unmarked ones within morphemes. What follows is also that longer clusters are more likely to be morphologically complex. The observation that morphotactic clusters are less constrained by phonological criteria is supported by the semiotic priority of morphology over phonology. In other words, the morpho-semantic motivation overrides the perceptual and articulatory difficulties posed by clusters.

The above hypotheses were tested in a series of studies on the lexica and corpora of several languages (cf. Zydorowicz et al., 2016²) as well as against the data from first (Dziubalska-Kołaczyk, 2019; Zydorowicz, 2010) and second language acquisition (Dziubalska-Kołaczyk & Zielińska, 2010, 2011; Dziubalska-Kołaczyk & Zydorowicz, 2014).

For example, of the 1734 cluster types in the Polish, there are only three types of 6-consonant clusters and 51 types of 5-consonant ones (see below Table 1). 100 percent of 6- and 5-member clusters and 95 percent of 4-member clusters are morphotactic.

¹ What is essential about NAD is that it does not refer to the syllable, but to the position of a cluster in a word. Preferability predictions are formulated separately for initial, medial and final clusters.

² The data for Polish and English have been collected for the Polish National Science Foundation (NCN) Project no. N N104 382540. Acknowledgements to my research team: Paulina Zydorowicz, Paula Orzechowska, Michał Jankowski, Piotr Wierzchoń, and Dawid Pietrala.

Table 1*Cluster Length in Polish*

Number of OC	Number of types
2	485
3	976
4	219
5	51
6	3
	1734

Also, morphologically complex clusters are largely dispreferred according to the phonotactic preferability criteria (measured by NAD). However, among the 14 most frequent initial clusters in the dictionary, four are dispreferred, namely /pɕ-, st-, sp-, sk-/. These call for additional explanations, e.g. /pɕ-/ occurs mostly in derivations involving the three highly productive prefixes, *przed*, *przy* and *prze*. It is clear that one needs to investigate the usage of clusters across corpora since highly dispreferred/marked clusters may achieve high frequencies, e.g. /dl-/, 3 occurrences in the dictionary, 172,698 in the corpus, mainly due to the word *dla* ‘for’; /kt-/, 7 occurrences in the dictionary, 92,997 in the corpus, due to *kto* ‘who’, *który* ‘which’, /gd-/, 3 occurrences in the dictionary, 75041 in the corpus, due to *gdy* ‘when’ and /zn-/, type frequency 19, token frequency 35424, due to *znać* ‘know’.

In English we looked at word final clusters: all CCCC clusters and almost all CCC clusters are morphonotactic. Similarly to Polish, the majority of morphologically complex clusters are phonologically dispreferred (according to NAD). However, again as in Polish, the impact of frequency of usage is often stronger than of phonological preferability, and marked clusters get high frequencies, e.g. final /-st, -nz, -ts/ alongside the unmarked /-nd, -nt/.

In German, we found 64 percent of all initial clusters to be NAD-preferred (Orzechowska & Wiese, 2015; Orzechowska & Dziubalska-Kołaczyk, in press).

The studies of first language acquisition (Yavaş & Marecka, 2013; Marecka & Dziubalska-Kołaczyk, 2014; Zydorowicz, 2010; Dziubalska-Kołaczyk, 2019) showed again an impact of frequency. For example, a Polish child (female, Poznań, age 1;7–3;2, a longitudinal study) produced /pɕ-/ most frequently, which is also the most frequent cluster in Polish, and the second most frequent cluster in her production was /st/, which is the third most frequent one in Polish. These results suggest (a) that frequency is not straightforwardly predictable from phonological preferability but rather represents an independent factor in its own right, and (b) that frequency (no matter what its causes may be) can even counteract and outweigh phonological

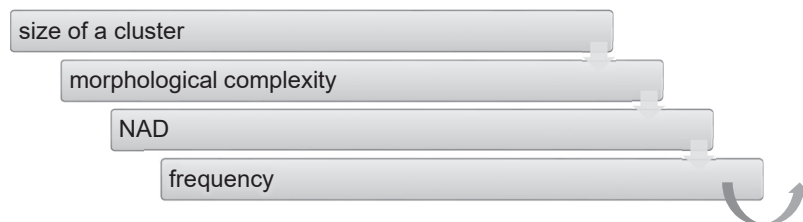
preferability when it comes to the ease and the speed with which consonant clusters are acquired. As Fikkert and Freitas (2004, p. 10) observed “it is important to consider the language system as a whole to interpret the data, both to explain differences between children acquiring the same language (i.e. the child’s own phonological system determines what optimal realizations for clusters are), and between children acquiring different languages”.

The observations concerning SLA have been informed by the data generated in our research on (mor)phonotactics (Dziubalska-Kołodziej & Zielińska, 2010, 2011; Dziubalska-Kołodziej and Zydorowicz, 2014; Marecka and Dziubalska-Kołodziej, 2014). We formulated and found support for the following predictions concerning L2/FL consonantal phonotactics and morphonotactics (see also Dziubalska-Kołodziej & Wrembel, in press for an analysis in terms of NGTA—Natural Growth Theory of Acquisition):

1. Clusters are difficult for learners so they modify and reduce them.
2. Cluster types which are common across languages (e.g., st-) seem to be easier despite their markedness, based on general observations.
3. Difficulty correlates with the universal phonotactic preferences: ‘good’ clusters are easier.
4. Less complex clusters (shorter) are less difficult.
5. Clusters are acquired in this order: medial > initial > final.
6. Dispreferred (marked) clusters are difficult for learners also when they are morphonotactic.
7. Children may learn morphonotactic clusters earlier.
8. Frequently used clusters are learned despite their markedness (corpus frequency overrides dictionary frequency).
9. Proficiency and metalinguistic awareness enhance the learning of clusters.

In conclusion, based on the numerous studies of lexicon, corpus, acquisition data as well as historical change and language processing (not exemplified here) of consonantal clusters we arrived at the following hierarchy of measures of cluster preferability: cluster size is on top, followed by morphological complexity, prior to phonological criteria of preferability, with frequency at the bottom. Crucially, however, frequency may override all the other criteria and end up on top of the hierarchy. This will not, however, make the most frequent cluster “unmarked” or the most natural one: it will simply be the most used one.

Figure 7
Hierarchy of Measures of Cluster Preferability



3. NATURAL COMPLEX SYSTEMS

The key to the convergence of complex systems and natural linguistics is to understand that the semiotic and functional ideas of natural linguistics are perceptual phenomena that arise from the underlying frequency profiles that emerge from the complex system of human speech. For instance, in the emergent environment of complex systems the traditional idea of a *vowel system* cannot apply as an essential phenomenon; as shown above, what people actually say does not come down to simple targets. The actual picture is much more complicated, with substantial overlaps between vowels in the *system*, and no central points that could be thought of as a *target*. Thus, the notion of language learning corresponding to nonlinear patterns of how each vowel may be realized cannot be said to be *correct*, but rather more *in an understandable range of what speakers do*. This is how perceptual ideas from natural linguistics can be associated with complex systems; what people perceive about the speech around them is subject to preferences, semiotics, and function.

Let us revisit the data on Polish consonant clusters presented above. If we apply a frequency analysis of the kind used in complex systems, we see that Polish consonant clusters do follow the A-curve distributional pattern (Figure 8). By “all” we mean the most frequent clusters studied in ALL word positions.

Figure 8
Polish Consonant Clusters in ALL Positions (Initial, Medial, Final)

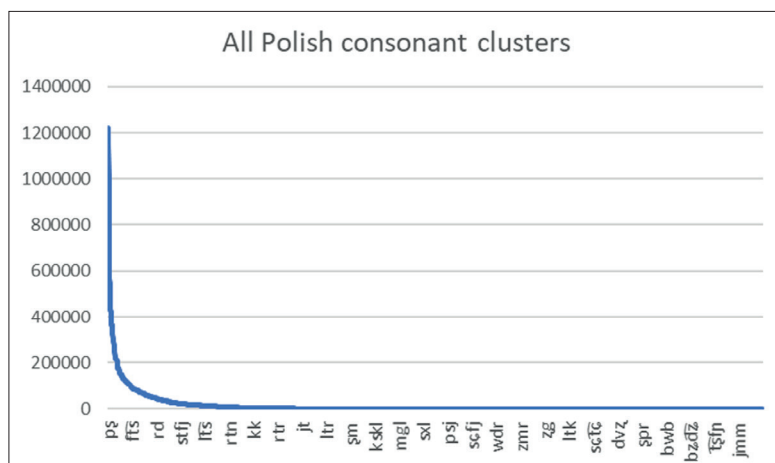


Table 2
Top 20 Most Frequent Polish Clusters

pos	len	cl	corpwtok
i	2	pɕ	1222231
m	2	st	1142579
i	2	pr	1005968
m	2	vj	752508
m	2	(tsj)	618007
m	2	(nts)	564556
m	2	nt	544810
m	2	(ete)	538049
i	2	st	442002
m	2	nd	430608
f	2	st	426348
m	2	ln	415876
m	2	(tʂn)	406234
i	2	kt	376132
i	2	mj	363071
i	2	sp	359765
i	2	vj	357597
m	2	ŋk	335182
m	2	dn	319117
m	2	mj	301749

The top twenty clusters all consist of two-consonant clusters. The first three-consonant cluster is in the fifty-first rank (medial \widehat{ksj}). The first four-consonant cluster does not occur until the one hundredth rank on the list (\widehat{jstf}). The frequency rankings thus support the idea that size of cluster is an important factor. The dispreferred clusters found frequently in the dictionary, /pɕ-, st-, sp-, sk-/, are all found in the top twenty ranks (st- and -st are both present). The cluster pɕ is one of two clusters with more than one million tokens in the corpus, with st- and -st not far behind at over 400,000 tokens apiece. This confirms that frequency is not predicable from preference, so preferences must be an independent factor.

In order to investigate preferences, let us first look at the frequency distribution of all word-initial clusters in a Polish corpus.

Figure 9

All Polish Initial Clusters

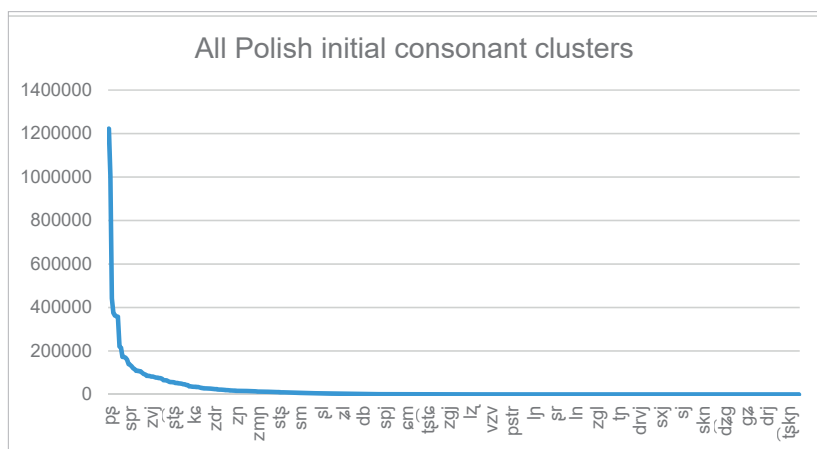


Table 3

Top 20 Most Frequent Initial Clusters With NAD

IPA transcription	Preferred cluster?	Lex
pɕV	No	Yes
prV	Yes	Yes
stV	No	
ktV	No	Yes
mjV	Yes	Yes
spV	No	
vjV	Yes	
krV	Yes	Yes

IPA transcription	Preferred cluster?	Lex
pjV	Yes	Yes
dIV	Yes	Yes
trV	Yes	Yes
grV	Yes	Yes
vwV	Yes	
skV	No	
sprV	Yes	
znV	No	
ɛfjV	Yes	Yes
drV	Yes	Yes
sfV	No	
brV	Yes	Yes

70% of the most frequent 20 clusters (between 1222231 and 108783 corpus frequency) are NAD-preferred, and the group includes only two 3C clusters. 12 of the 20 clusters are purely lexical (phonotactic), the remaining ones are morphologically complex, i.e. they are morphonotactic or mixed. Among the 12 phonotactic clusters, only pʂV and ktV are NAD-dispreferred. This data shows that, firstly, size of a cluster is at the top of the hierarchy of measures; secondly, morphological complexity overrides phonotactic preferences; thirdly, phonotactic preferences work at the level of 70% for the whole group of the top 20 clusters, and at the level of 83% for phonotactic subgroup. Finally, frequency of usage motivates the remaining clusters, like pʂV and ktV, already discussed earlier in the paper. The high percentage of NAD-preferred clusters among the top-ranked clusters on the A-curve suggests that the idea of preference is strongly aligned with frequency. That is, our analytical idea of preferences comes from top ranked, high frequency clusters as they occur in Polish. Several dispreferred clusters do exist in the top ranks along the A-curve, so preference seems not to be driving frequency, and at the same time the high percentage of preferred top ranked types drives our perception of preference.

Table 4*All Initial Clusters With NAD*

IPA transcription	Preferred cluster?	Lexical
pʂV	No	Yes
prV	Yes	Yes
stV	No	
ktV	No	Yes
mjV	Yes	Yes
spV	No	
vjV	Yes	
krV	Yes	Yes
pjV	Yes	Yes
dIv	Yes	Yes
trV	Yes	Yes
grV	Yes	Yes
vwV	Yes	
skV	No	
sprV	Yes	
znV	No	
cfjV	Yes	
drV	Yes	Yes
sfV	No	
brV	Yes	Yes
strV	Yes	
fʂV	No	
zdV	No	
tʂV	No	Yes
gdV	No	Yes
erV	Yes	Yes
swV	Yes	
gwV	Yes	Yes
zvjV	Yes	
fspV	No	
spʂV	No	
bjV	Yes	Yes
dvV	No	Yes

IPA transcription	Preferred cluster?	Lexical
zmjV	Yes	
kjV	Yes	Yes
klV	Yes	Yes
plV	Yes	Yes
zwV	Yes	
kfV	No	Yes
fʃ̥V	No	
zvV	No	
vrV	Yes	
ʃʃ̥V	No	
frV	Yes	
dɲV	Yes	
pwV	Yes	Yes
zgV	No	
zrV	Yes	
tʃwV	Yes	Yes
mɲV	No	Yes
ʃkV	No	Yes
vzrV	Yes	
skwV	Yes	
dwV	Yes	Yes
zbV	No	
fprV	Yes	
keV	No	Yes
vɲV	Yes	
giV	Yes	Yes
xʃ̥V	No	Yes
tfV	No	Yes
fpwV	Yes	
ftV	No	
elV	Yes	Yes
blV	Yes	Yes
ʃtV	No	Yes
fskV	No	
kʃ̥V	No	Yes

IPA transcription	Preferred cluster?	Lexical
mwV	Yes	
gdzV	No	Yes
zdrV	Yes	
zvrV	Yes	
tfjV	Yes	Yes
gmV	Yes	Yes
stfjV	No	
trfV	No	Yes
tʃtV	No	Yes
xteV	No	Yes
zmV	Yes	
emjV	No	
vzglV	No	
zjV	Yes	
vzV	No	
ʃfV	No	Yes
zɲV	No	
stsV	No	
zbjV	Yes	
skrV	Yes	
ferV	Yes	
eteV	No	
xfV	No	Yes
gvV	No	Yes
twV	Yes	Yes
tʃfV	No	Yes
gzV	No	Yes
kfjV	Yes	Yes
bwV	Yes	Yes
zgwV	Yes	
zmpV	No	
zvwV	Yes	
zlV	Yes	
fpV	No	
psV	No	

IPA transcription	Preferred cluster?	Lexical
kwV	Yes	Yes
ʂpV	No	Yes
vzV	No	
dvjV	Yes	Yes
fsxV	No	
zrV	Yes	Yes
fstV	No	
stfV	No	
glV	Yes	Yes
stʂV	No	
zgrV	Yes	

In the extended search (the table above), 53% of the 114 clusters along the ascending curve (between 1222231 and 10168) are NAD-preferred. Note that the 114th ranked cluster is two orders of magnitude (100 times) less frequent than the 1st rank. Among these, there are two 4C clusters, so, the measure of size works again. The decreased percentage of the NAD-preferred clusters with decreased frequency was also expected. 54 among those 114 clusters are lexical: of these 54, 31 are NAD-preferred, i.e. 57%. Thus, again, the preferability is slightly higher among the purely lexical clusters. The decreased percentage of NAD-preferred clusters in the longer list shows that the preference is more associated with the most frequent, top ranked forms and declines with the decline in frequency.

Let us now examine all final clusters (209) of Polish.

First of all, there are less than half as many final clusters than initial ones and they are less frequent than the initial clusters: the top ranked cluster Vst (426348) has 3 times fewer tokens than the top ranked initial pʂV (1222231). Secondly, only four of the top 20 clusters overlap with the initial 20 ones (st, kt, tr, sk) which clearly shows different preferences for initial and final position. As to NAD (see Table 5 below), only 40% of the top 20 clusters are preferred. This, however, is predictable by the order of preference with regard to position: medial > initial > final. Six of the 20 clusters can be morphologically complex, i.e. they can contain a consonantal suffix (+tɛ or +w), which additionally accounts for their markedness.

Figure 10
Most Frequent of All Final Clusters in Polish

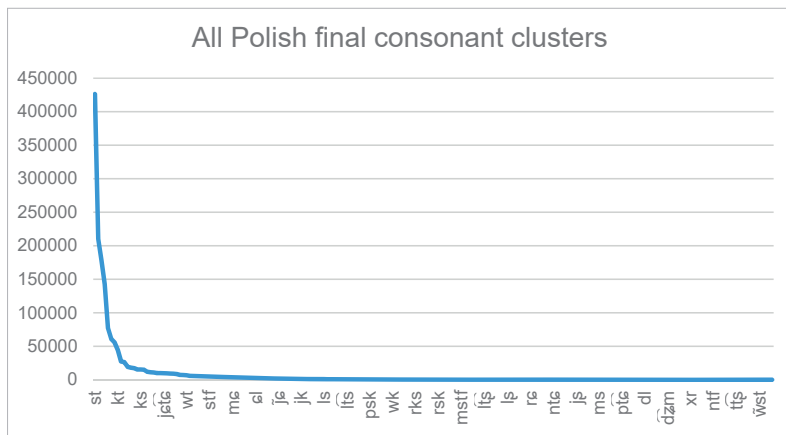


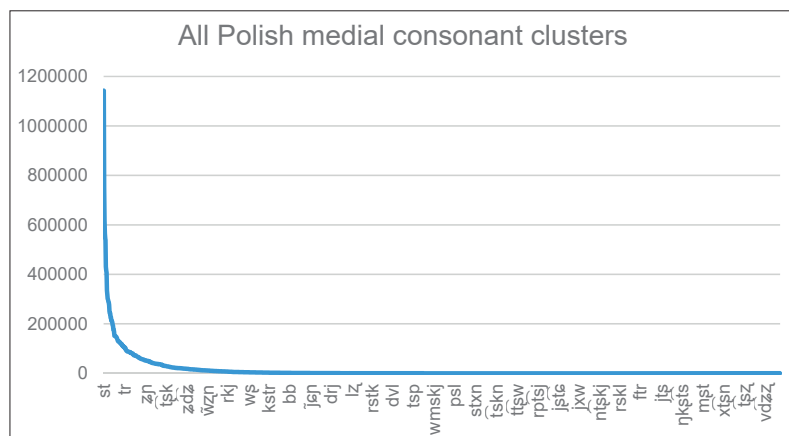
Table 5
Top 20 Most Frequent Final Clusters With NAD

IPA transcription	Preferred cluster?
Vst	No
Vete	No
Vnts	No
Vjete	No
Vkt	No
Vsk	No
Vsw	No
Vdw	No
Vks	No
Vtr	No
Vgw	No
Vns	No
Vnt	Yes
Vŋk	Yes
Vrt	Yes
Vrs	Yes
Vrm	Yes
Vjete	Yes
Vrk	Yes
Vw̃ŋ	Yes

We turn now to the Polish medial clusters (1902).

Figure 11

Most Frequent of All Medial Clusters in Polish



The number of medial clusters itself, i.e. 1902, in comparison to the number of initial (457) and medial (209) clusters, shows the overwhelming preference for a cluster to be situated word medially. Frequency of the top ranked medial cluster is comparable to the top initial one: VstV (1142579). Nine of the top 20 medial clusters do not overlap with either initial or final top 20, which again proves that different preference criteria apply to each position in a word. 60% of the top 20 are NAD-preferred (cf. Table 6 below). With reference to morphological complexity, medial clusters have a great potential to be derived, due to the richness of affixes in Polish. Across the three positions, then, NAD preferences are aligned with frequency in the top twenty ranks of the A-curve at rates of 70% (initial), 60% (medial), and 40% (final). The global idea of preferences is strongly supported by these high rates, while at the same time it is clear that preferences apply differently in the different positions. Preferences in initial and medial positions are predicted to be at higher frequencies than those in final position according to the NAD/NL framework.

Table 6*Top 20 Most Frequent Medial Clusters With NAD*

IPA transcription	Preferred cluster?
VstV	Yes
VntsV	Yes
VntV	Yes
VɛtɛV	Yes
VndV	Yes
VlnV	Yes
VtʂnV	Yes
VŋkV	Yes
VtkV	Yes
VktV	Yes
VskV	Yes
VvnV	Yes
VvjV	No
VtsjV	No
VdnV	No
VmjV	No
VdɲV	No
VrtV	No
VrmV	No
VvɲV	No

Figures 12, 13 and 14 show the A-curves and top twenty lists just for Polish three-consonant clusters (not all size clusters) in initial, medial, and final position, respectively.

Figure 12
Polish Initial Three-Consonant Clusters

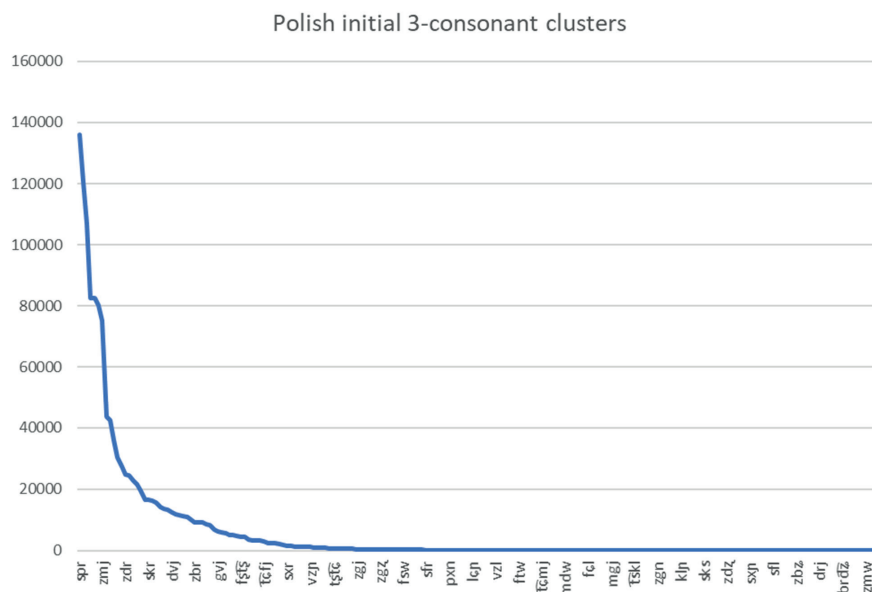


Table 7
Polish Initial Three-Consonant Clusters

spr	136070
ɛfj	120651
str	107047
zvj	82604
fsp	82576
spɕ	80082
zmj	75361
vzr	43653
skw	42476
fpr	35973
fpw	30218
fsk	27467
zdr	24856
zvr	24403
tfj	22824

trf	21651
emj	19545
zbj	16467
skr	16388
fer	16332

\widehat{jete}	18081
$jete$	9854
\widehat{jsts}	5102
stf	4949
st \mathring{s}	4211
kst	4049
nt \mathring{s}	3723
$\widehat{j}sk$	3294
$\widehat{t}stf$	3105
ηkt	2499
$\widetilde{w}sk$	1983
str	1419
jsk	1400
psk	525
wst	396
rks	304
rsk	227
\widehat{rete}	160
rst	145
$\widehat{r\mathring{s}t\mathring{s}}$	103

Figure 13
Polish Medial Three-Consonant Clusters

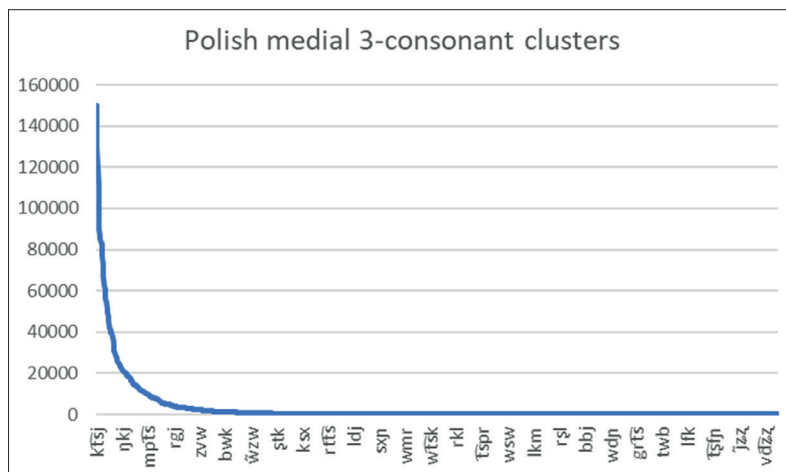
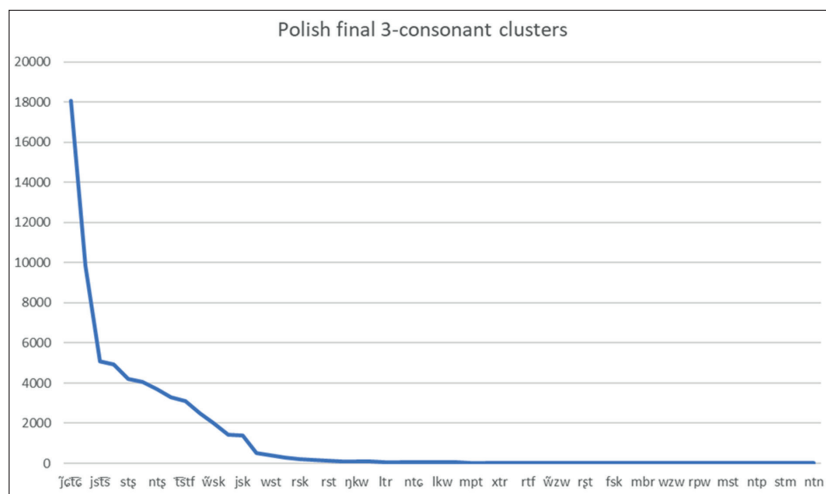


Table 8
Polish Medial Three-Consonant Clusters

ktsj	149828
nsk	129016
str	120285
nts	111041
fsk	89672
nkʂ	86650
tst	85303
stk	83620
ntr	82371
ʃsk	76804
rfʂ	72855
spj	67024
tstf	62562
ntk	60867
jsk	56055
ʃsk	54506
stʂ	51062
jsts	49230
tskj	47577
nst	45457

Figure 14*Polish Final Three-Consonant Clusters***Table 9***Polish Final Three-Consonant Clusters*

j̃eṭe	18081
jst̃s	9854
st̃s	5102
nt̃s	4949
t̃stf	4211
w̃sk	4049
jsk	3294
wst	3723
rsk	3105
rst	2499
ŋkw	1983
ltr	1419
nte	1400
lk̃w	525
mpt	396
xtr	304
rtf	227
w̃zw	160
r̃st	145
fsk	103
mbr	
w̃zw	
rp̃w	
mst	
ntp	
stm	
ntn	

While Figures 9, 10 and 11 all show clear A-curves, the order of clusters is again quite different. The two top-ranked initial clusters do not appear in the top twenty ranks of the medial or final clusters; the third ranked initial and medial cluster, *str-*, does not appear until the twelfth rank among final clusters. The same is true for two consonant clusters and for clusters larger than three, although we do not show the charts and lists here. Thus, it is clear that preferences among clusters in all three positions are different, as NAD predicts. These preferences in different positions are created by the separate frequency profiles for each position for clusters, not by the overall pattern of frequency.

CONCLUDING REMARKS

We have tried to demonstrate that the preferences based on cognitive, semiotic and functional principles in natural linguistics have a complicated relationship with frequency patterns as they arise from the complex system of speech. The hypothesis is that the A-curves generate the perceptions that are described by NL, rather than the other way around. The distribution of the data (in A-curves) should yield preferences, at every level of scale, for certain clusters. Some dispreferred clusters are found among the top ranks at every level of scale, but at the same time the balance of types at every level of scale gives rise to the perception of preferences. The underlying A-curves from the complex system are always present, at every level of analysis (all clusters, initial clusters, final clusters, clusters of different sizes). What preferences speakers develop depends essentially on these frequency patterns, although speakers' preferences do not rule out exceptional frequencies for some clusters. The A-curves allow for exceptional clusters, like *pɛ-*, to emerge even while the general trend of the top ranks gives the preferences described in NL. Similarly, morphological complexity and NAD are informed, but not determined absolutely, by the underlying frequency patterns. It is clear that preferences and regularities are not absolute: they are tendencies rather than universals. This means that a knowledge of complex systems can enhance natural linguistics by showing that it is not merely a logical theory but one that arises from actual linguistic evidence as Dressler and others have claimed in multiple publications.

REFERENCES

- Burkette, A. (2016). *Language and material culture: Complex systems in human behavior*. John Benjamins.
- Burkette, A., & Kretzschmar Jr., W. A. (2018). *Exploring linguistic science: Language use, complexity and interaction*. Cambridge University Press.
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
- Bybee, J. (2002). Sequentiality as the basis for constituent structure. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of prelanguage* (pp. 109–132). John Benjamins.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Donegan, P. J. (1978/1985). *On the natural phonology of vowels*. Garland.
- Donegan, P., & Stampe, D. (1979). The study of natural phonology. In D. A. Dinnsen (Ed.), *Current approaches to phonological theory* (pp. 126–173). IUP.
- Donegan, P., & Stampe, D. (2009). Hypotheses of natural phonology. *Poznań Studies in Contemporary Linguistics*, 45(1), 3–31.
- Dressler, W. U. (1985). Explaining natural phonology. In C. Ewen & J. Anderson (Eds.), *Phonology yearbook 1* (pp. 29–50). Cambridge University Press.
- Dressler, W. U. (1996). Principles of naturalness in phonology and across components. In B. Hurch & R. A. Rhodes (Eds.), *Natural phonology: The state of the art* (pp. 41–52). De Gruyter.
- Dressler, W. U. (2011). The rise of complexity in inflectional morphology. *Poznań Studies in Contemporary Linguistics*, 47, 159–176.
- Dressler, W. U., & Dziubalska-Kołaczyk, K. (1994). Functional analysis in the study of second language acquisition. *Functions of Language*, 1(2), 201–228.
- Dressler, W. U., & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics. *Rivista di Linguistica*, 18(2), 249–266.
- Dziubalska-Kołaczyk, K. (Ed.). (2001). *Constraints and preferences. Trends in linguistics*. Studies and monographs 134. De Gruyter.
- Dziubalska-Kołaczyk, K. (2002). Challenges for natural linguistics in the twenty first century: A personal view. *University of Hawaii Working Papers in Linguistics: Vol. 23. (2001–2002)* (pp. 15–39). University of Hawaii at Mānoa.
- Dziubalska-Kołaczyk, K. (2006). Modern natural phonology: The theory for the future. In J. Fisiak (Ed.), *English language, literature and culture. Selected papers from the 13th PASE conference, Poznań 2004* (pp. 1–10). Uni-Druk.
- Dziubalska-Kołaczyk, K. (2012). Modern natural phonology and phonetics. In E. Cyran, H. Kardela, & B. Szymanek (Eds.), *Sound, structure and sense. Studies in memory of Edmund Gussmann* (pp. 199–210). Wydawnictwo KUL.
- Dziubalska-Kołaczyk, K. (2014). Explaining phonotactics using NAD. *Language Sciences*, 46(s), 6–17.
- Dziubalska-Kołaczyk, K. (2015). Are frequent, early and easy clusters also unmarked? *Italian Journal of Linguistics*, 27(1), 29–44.
- Dziubalska-Kołaczyk, K. (2019). On the structure, survival and change of consonant clusters. *Folia Linguistica Historica*, 40(1), 107–127.
- Dziubalska-Kołaczyk, K., & Weckwerth, J. (Eds.). (2008). *Future challenges for natural linguistics* (2nd ed.). Lincom.
- Dziubalska-Kołaczyk, K., & Wrembel, M. (in press). Natural growth theory of acquisition (NGTA): evidence from (mor)phonotactics. In V. Sardegna & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training*. Springer.
- Dziubalska-Kołaczyk, K., & Zielińska, D. (2010). Predicting phonotactic difficulty in second language acquisition. In A. S. Rauber, M. A. Watkins, R. Silveira, & R. D. Koerich (Eds.), *The acquisition*

- of second language speech: Studies in honor of Professor Barbara O. Baptista* (pp. 281–304). Editora Insular.
- Dziubalska-Kołodziej, K., & Zielińska, D. (2011). Universal phonotactic and morphotactic preferences in second language acquisition. In K. Dziubalska-Kołodziej, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (pp. 53–64). Peter Lang Verlag.
- Dziubalska-Kołodziej, K., & Zydorowicz, P. (2014). The production of high-frequency clusters by native and non-native users of Polish. *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia working papers in applied linguistics, 5, COPAL* (pp. 130–144).
- Ellis, N., & Larsen-Freeman, D. (Eds.). (2009). *Language as a complex adaptive system*. Wiley-Blackwell.
- Fikkert, P., & João Freitas, M. (2004). The role of language-specific phonotactics in the acquisition of onset clusters. In L. Cornips & J. Doetjes (Eds.), *Linguistics in the Netherlands 2004* (pp. 1–12). John Benjamins.
- Gould, S. J. (2003). *The hedgehog, the fox, and the magister's pox: Mending the gap between science and the humanities*. Three Rivers.
- Hawking, S., & Mlodinow, L. (2010). *The grand design*. Bantam.
- Hiver, P., & Al-Hoorie, A. H. (2020). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.
- Holland, J. (1998). *Emergence: From chaos to order*. Basic.
- Hopper, P. (1987). Emergent grammar. *Berkeley Linguistics Society, 13*, 139–157.
- Kretzschmar Jr., W. A. (2009). *The linguistics of speech*. Cambridge University Press.
- Kretzschmar Jr., W. A. (2015) *Language and complex systems*. Cambridge University Press.
- Langacker, R. (1988). A usage based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). John Benjamins.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141–165.
- Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). De Gruyter.
- Maddieson, I. (1999) 'In search of universals', in J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.) *Proceedings of the 14th International Congress of Phonetic Sciences 1999*, (Vol. 3, pp. 2521–2528).
- Maddieson, I. (2009). Calculating phonological complexity. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 85–109). De Gruyter.
- Mandelbrot, B. (1968). Information theory and psycholinguistics. In R. C. Oldfield & J. C. Marshall (Eds.), *Language: Selected readings* (pp. 263–275). Penguin.
- Mandelbrot, B. (1982). *The fractal geometry of nature*. Freeman.
- Marecka, M. & Dziubalska-Kołodziej, K. (2014). Evaluating the models of phonotactic constraints on the basis of the sC cluster acquisition data. *Language Sciences, 46*, 37–47.
- Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.
- Ohala, J. J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. The parasession on the syllable in phonetics and phonology. *CLS, 26*(2), 319–338.
- Olsen, R. M., Olsen, M., Stanley, J. A., Renwick, M. E. L., & Kretzschmar Jr., W. A. (2017). Methods for transcription and forced alignment of a legacy speech corpus. *Proceedings of Meetings on Acoustics, 30*(1), 060001. <https://doi.org/10.1121/2.0000559>
- Orzechowska, P., & Wiese, R. (2015). Preferences and variation in word-initial phono-tactics: A multidimensional evaluation of German and Polish. *Folia Linguistica, 49*(2), 439–486.

- Orzechowska, P., & Dziubalska-Kołaczyk, K. (in press). Gradient phonotactics and frequency: A study of German initial clusters. *Italian Journal of Linguistics*.
- Peirce, C. S. (1965). *Collected Papers* (Charles Hartshorne & Paul Weiss, Eds.). Harvard University Press.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos. Man's new dialogue with nature*. New York: Bantam Books.
- Stampe, D. (1969). The acquisition of phonetic representation. *CLS*, 5, 443–453.
- Stampe, D. (1973/1979). *A dissertation on natural phonology*. IULC.
- Yavaş, M., & Marecka, M. (2013). Acquisition of Polish #sc clusters in typically-developing children and in children with phonological disorders. *International Journal of Speech-Language Pathology*, 16(2), 132–141.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.
- Zydorowicz, P. (2010). Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *Poznań Studies in Contemporary Linguistics*, 46(4), 565–588.
- Zydorowicz, P., Orzechowska, P., Jankowski, M., Dziubalska-Kołaczyk, K., Wierchoń, P., & Pietrala, D. (2016). *Phonotactics and morphonotactics of Polish and English. Theory, description, tools and applications*. Adam Mickiewicz University Press.

COMPLEX NATURAL SYSTEMS FOR LANGUAGE

Summary

In this paper we want to show how complex systems theory can inform the theoretical development in natural phonology. This development leads us to propose complex natural systems modelling for language. We explain how complex systems help us understand both theoretical concepts from natural phonology, for example the concept of preference, as well as variation in language data, for example in English vowel realizations and Polish consonant clusters. We select from our respective research areas those exemplifications which clearly demonstrate substantial variation in language use.

Keywords: complex systems; natural linguistics; sociolinguistics; language acquisition; phonotactics.

ZŁOŻONE SYSTEMY NATURALNE DLA JĘZYKA

Streszczenie

W artykule tym chcemy pokazać, w jaki sposób teoria systemów złożonych może wpłynąć na rozwój teoretyczny fonologii naturalnej. Rozwój ten prowadzi nas do zaproponowania modelowania złożonych systemów naturalnych dla języka. Wyjaśniamy, w jaki sposób systemy złożone pomagają nam zrozumieć zarówno koncepcje teoretyczne z fonologii naturalnej, na przykład koncepcję preferencji, jak i wariantywność danych językowych, na przykład w angielskich realizacjach samogłosek i polskich zbitkach spółgłoskowych. Wybieramy z naszych obszarów badawczych te przykłady, które wyraźnie pokazują znaczną wariantywność w użyciu języka.

Słowa kluczowe: systemy złożone; językoznawstwo naturalne; socjolingwistyka; akwizycja języka; fonotaktyka.

KATARZYNA DZIUBALSKA-KOLACZYK is Full Professor and Vice-Rector for research of Adam Mickiewicz University in Poznań. She was Dean of the Faculty of English. She has published extensively (ca. 160 publications) on phonology, phonetics and language acquisition. In her works she has been pursuing and advocating the Natural Linguistic approach to language. Her books include *A Theory of Second Language Acquisition within the Framework of Natural Phonology*, *Beats-and-Binding Phonology*. Her recent co-authored book is *Phonotactics and Morphonotactics of Polish and English*. She is Editor-in-Chief of *Poznań Studies in Contemporary Linguistics* and organizes *Poznań Linguistic Meetings*. She was a Senior Fulbright scholar in 2001–2002 (University of Hawai'i at Manoa) and a visiting scholar at the University of Vienna (1991–94, 1998). She is a member of Academia Europaea, Agder Academy, Linguistic Committee of PAN, a corresponding member abroad of the Austrian Academy of Sciences. In 2013–2014, she was President of Societas Linguistica Europaea. She has supervised 22 PhD dissertations.

WILLIAM A. KRETZSCHMAR, JR. teaches English as Willson Professor in Humanities at the University of Georgia. He also has an appointment at the University of Oulu. His major publications include *Exploring Linguistic Science*, *The Emergence and Development of English*, *The Routledge Dictionary of Pronunciation for Current English*, *Language and Complex Systems*, *The Linguistic of Speech*, *The Oxford Dictionary of Pronunciation for Current English*, and *The Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. He edited the Linguistic Atlas Project, a large American dialect project, for 34 years. He is also active in corpus linguistics, where he directed corpus and text encoding activities for a National Cancer Institute grant to study tobacco documents. He has been influential in development of digital methods for analysis and presentation of language variation, including the application of complexity science.