Form-Meaning Regularities in Old English Lexicon

Nataliia Drozhashchikh · Elena Efimova · Evgenia Meshcheryakova

Abstract. This article deals with the form-meaning hypothesis in Old English within the theory of arbitrariness/non-arbitrariness. It focuses on the relations between the initial grapheme (phoneme) in a word and its lexical semantics and aims to reveal any non-arbitrary form-meaning associations at the lexical level. The data include <w>-, <s>-, <h>-, and <p>-words from a Thesaurus of Old English. The methodology employs statistical methods (Chi-square test, the coefficient of contingency, the contributions to the Chi-square) within Python realization. Our primary hypothesis is that alliteration—regular repetition of onsets in Old English lexemes, could stand for the regularities in the semantics of these words. We extrapolate the initial research and underlying hypothesis to lexical data in general. The findings demonstrate non-arbitrary form-meaning regularities at the level of the entire Old English lexicon—the tendency of words sharing initial graphemes to be attracted to certain semantic categories.

1. Introduction

The present study focuses on form-meaning relationships in the Old English lexicon. Form-meaning mappings have long been in the focus of attention in semiotics and linguistics. They were first mentioned in the theory of the correctness of names in Plato's *Cratylus*. According to *phusei* approach, it was proposed that "there is a kind of inherent correctness in names" and according to *sunthēkē and homologia* approach, it

Nataliia Drozhashchikh 0000-0002-5910-2402
Tyumen State University, Volodarskogo 6, 625003 Tyumen, Russia
E-mail: n.v.drozhashhikh@utmn.ru

Elena Efimova 0000-0002-6584-965X
Tyumen State University, Volodarskogo 6, 625003 Tyumen, Russia
E-mail: elena.efimova@gmail.com

Evgenia Meshcheryakova 0000-0001-8748-640X
Independent researcher
E-mail: evg.meshch@gmail.com

Y. Haralambous (Ed.), *Graphonlinguistics in the 21st Century 2020. Proceedings*
Graphonlinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 5.
Fluxus Editions, Brest, 2021, pp. 739–754. https://doi.org/10.36824/2020-graf-droz
ISBN: 978-2-9570549-7-8, e-ISBN: 978-2-9570549-9-2

was believed that there is no "correctness of names other than convention and agreement".[1] The phusei approach has served as the basis for developing the theory of non-arbitrariness (iconicity) that includes into its scope all motivated form-meaning cases (onomatopoeia, sound symbolism, phonaesthemes, ideophones, etc.). In recent decades an increasing number of publications on non-arbitrary coding in language have seemed to undermine the thesis of linguistic conventionality and have made it clear that there cannot be a place for "convention and agreement" dogma in linguistic theory. While studying frequency and complexity of letters and script (Altmann, 2004); universal lexical semantics across languages (Blasi et al., 2016; Youn et al., 2016); phonological systematicity (Monaghan, Shillcock, Christiansen, and Kirby, 2014); correlation of orthographic/phonological form (Jee, Tamariz, and Shillcock, 2018); form-meaning mapping in alliterative verse (Cornell, 1981); semantic functions of phonemic clusters (Lvova, 2005); the semantics of graphemes (Slaměníková, 2019), modern scholars provide evidence that language combines both arbitrary and non-arbitrary relations.

Form-meaning relations in earlier stages of language have not been sufficiently studied. A few papers focus directly on this kind of relationship in Old English, in particular, on sound symbolism, submorphemic iconicity, and form-meaning association in alliteration (Cornell, 1981; Jespersen, 1922; Minkova, 2003; Philps, 2008; C. A. Smith, 2016). While studying the relations between the phonemes and the semantic classes like "sound," "tone," "size," "movement," "human body," etc., the authors ascribe the meaning to the phonemic clusters themselves. E.g., Philps (2008) claims that word-initial phonaesthemes "are endowed with a potential for meaning". Cornell (1981) does not attribute meaning to the phonemes but reveals the tendency of alliterating sounds to be connected with certain connotative meanings. As far as the entire Old English lexicon is concerned, the research pertaining to the association of the initial phoneme in the words and their referential meaning is insufficient. Meanwhile, referentiality is directly related to the process of non-arbitrary coding in older languages as early nominations are more iconic (see Atkinson, Mills, and Smith, 2019).

Historical linguistics and the theory of evolution put an important emphasis on the role of non-arbitrary coding in language formation and development. Older languages represent the phase of development, where a clear motivation of the nomination and word formation processes are possible, and the original etymological basis of referential meanings has not yet been suppressed. The formation processes that take place in different parts of the language system at the earlier stages of its development provide more information about possible form-meaning relationships than in modern languages. For example, it

1. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0172

is known that as the vocabulary grows more abstract, the original meaning of the word, which, as a rule, is concrete, becomes obscure. The growing number of derivatives makes it difficult to extract monomorphic lexemes that retain their original non-derivational meanings. The penetration and assimilation of lexical borrowings in the receiving language complicate the distinction between the original and borrowed vocabulary, which does not help to understand the processes taking place in the lexicon. All these processes violate the original linguistic systematicity in order to become the source of new systematic relations in the further stages of linguistic development.

In this paper, we attempt to test the non-arbitrary form-meaning hypothesis in the Old English lexicon. Statistical methods and semantically (conceptually) organized dataset of A Thesaurus of Old English allowed us to test the form-meaning association at the lexical level and conclude with certain assumptions a statistically significant form-meaning correlation.

The structure of the article is the following. In Introduction, we propose a hypothesis and set the objectives of the study; in Sections 2-4 we present the outline of the previous work and describe basic terminology; in Section 5 we outline the data and methods of the research and summarize the obtained results in Section 6. In Section 7 we outline the possible prospects of the current research and reveal its limitations.

2. Old English Lexicon

Old English is "the language spoken by the Germanic inhabitants of Britain" (5th–11th c.) in which prosaic and poetic texts were written (Fulk, 2014) and one of the earliest periods of language development (7th–11th c.). Old English prosaic texts include the translations of the Bible, Gospels, Psalter, Wulfstan's Homilies, Ælfric's works (religious texts); law codes, wills, and charters (legal texts); The Anglo-Saxon Chronicle (documentary prose); King Alfred's original compositions and translations from Latin (literary, philosophical, and didactic prose), historical works, and medical tracts. Poetry is represented by heroic and elegiac poems, religious and lyrical texts, magical and didactic poems, and riddles. Alliterative poetic texts, in particular, Beowulf, Genesis, Exodus, Cynewulf's poems Elene, Juliana, Andreas; Judith, The Dream of the Rood, The Wanderer, The Seafarer, Metrical Charms, and others have gained an important place in the history of Old English (Godden, 1992). There is a fairly large number of the surviving Old English texts. Nevertheless, religious texts are thematically dominant, which may impose constraints on the further analysis of the lexis.

The Old English lexicon, presented in A Thesaurus of Old English, was collected by lexicographers from the English texts of 7th–11th c. and

contains lexical layers of different chronological and etymological depth (Pollington, 1993). The lexicon includes the vocabulary of Latin/Greek origin, neutral groups of words of different genres, and a small amount of colloquial vocabulary (this layer of vocabulary cannot be fully selected due to the lack of speech fixation). Along with the vocabulary of different registers, the Old English lexicon includes purely poetic vocabulary found mainly in poetry and nowhere else (Barney, 1985).

The Old English lexicon and its various linguistic and stylistic aspects have been studied by many scholars (Kastovsky, 1992; J. J. Smith, 2009). Nevertheless, we are not aware of the works that study non-arbitrary form-meaning relationships in the Old English lexicon as a whole.

3. Form-Meaning Relationships: Basic Terminology

Within the framework of non-arbitrary relationships in language, there are a number of terms that constitute the conceptual basis for this research. They include the terms linguistic sign, signifier, signified, lexeme, initial sound, grapheme, referential meaning, arbitrariness, iconicity, similarity, isomorphism, analogy, systematicity.

Linguistic signs represented by words or lexemes—the major units of vocabulary, are identified by the two components—the form (sounds/graphemes) and the content (meanings). In Saussure's theory these components are signifier and signified. Signifier is represented as a sound or graphical form of a word. The most relevant component of the form is an initial sound (phoneme). It is the smallest structural unit of language that carries important information since it is cognitively and positionally marked. An initial phoneme is connected with an initial grapheme—an orthographic representation of a sound. Signified is a word meaning—a quantum of sense revealing information about entities, processes, ideas, events in the world. Since there are a number of word meanings (lexical, grammatical, social, connotative, pragmatic, etc.) we discuss only conceptual meanings. They differ from other forms of meanings in that they refer to a cognitive content of a word. We can consider a conceptual meaning as the referential meaning of a word—an entry that is usually given in a dictionary.

'Signifier/signified' relations can be arbitrary or non-arbitrary (iconic). Arbitrariness does not hold any direct natural connection between the sign and its meaning and can be applied to the majority of linguistic signs. The principle of "whatever name you give to a thing is its right name" (Plato) became the fundamental principle of language whereby linguistic signs are considered arbitrary or conventional (F. de Saussure). Non-arbitrariness or iconicity involves the relations of similarity, isomorphism or analogy between some aspects of form and meaning.

The three terms are nearly identical in meaning, with similarity being a "one-to-one mapping of a Euclidean space onto itself"[2], isomorphism is "a correspondence (relation) between objects or systems of objects expressing the equality of their structures in some sense"[3] (cf. Givón (1985)), and "analogy between S and T is a one-to-one mapping between objects, properties, relations and functions in S and those in T".[4] Non-arbitrariness involving these relations enables us to predict word meanings. E.g., a word form (sounds/graphemes in *moo*) is associated with the referent (a mooing cow) and bears some resemblance to its semantics ('the sound produced by a cow').

In recent years, the term systematicity has been increasingly used. It is understood as the statistical form-meaning regularities found in "localized form-meaning patterns" or "across the lexicon as a whole" (Gutiérrez, Levy, and Bergen, 2016, p. 2379). The content and status of the term has not yet been clearly defined since it is applied both to arbitrary and non-arbitrary form-meaning relationships. It is explained as arbitrary statistically regular patterning of sounds (Dingemanse et al., 2015) or as "strong, non-negligible lexicon-wide non-arbitrariness" (Gutiérrez, Levy, and Bergen, 2016, p. 2380). The 'information' nature of the term is identified in Pimentel et al. (2019, p. 1752) that estimates the mutual information between the form and the meaning of a linguistic sign, i.e., the word-form/semantic distance/similarity. The authors point out that systematicity can be understood more broadly—as an umbrella term for all cases of regular patterning in language. In that case, systematic relations are manifested in obligatory grammatical/semantic oppositions, e.g., in grammatical categories of case/number or semantic oppositions of hypernymy/hyponymy.

4. Non-Arbitrary Form-Meaning Relationships

Signifier/signified non-arbitrary relationships are studied in the theory of non-arbitrariness (iconicity). The research in the field of iconicity is extensive. Iconic and motivated signs are explored in language acquisition (Winter, Perlman, Perry, and Lupyan, 2017), cognitive (Wilcox, 2004) and neurolinguistic studies (Aryani, Jacobs, and Conrad, 2013); (Perniss and Vigliocco, 2014); (Monaghan, Shillcock, Christiansen, and Kirby, 2014), poetics (Tsur, 2002), etc. More and more studies focus on the evolutionary aspects of iconicity (Zlatev, Żywiczyński, and Wacewicz, 2020). In this and similar research, the authors define non-arbitrary form-meaning relationships, develop their classifications, and

2. http://encyclopediaofmath.org/index.php?title=Similarity&;oldid=31636
3. http://encyclopediaofmath.org/index.php?title=Isomorphism&oldid=21572
4. https://plato.stanford.edu/archives/spr2019/entries/reasoning-analogy/

disclose their nature. The terms (correspondence, equality, resemblance, congruence, equivalence, identity, analogy) that are used to interpret them are numerous, each reflecting subtle aspects of form-meaning association (mapping, correlation).

The definitions/mechanisms of non-arbitrary relationships vary depending on different types of linguistic signs. In onomatopoeia, non-arbitrary relationships are defined as the relations of low/high-order similarity or resemblance and structural similarity between form and meaning (Winter, Perlman, Perry, and Lupyan, 2017); (Dingemanse et al., 2015). In the former case, physical sensoriperceptual properties of the referent (usually emitted sounds) are imitated in the perceptual/graphical properties of the word form (sounds/graphemes) and are associated with some features of the lexical meaning of the word. The imitation of sensory properties refers to the low-order similarity where the attributes of the objects are compared. In the latter case, there is a relational similarity in the elements of structure (high-order similarity), where the relations between objects are compared, e.g., the sequential order of the events imitates the word order in a sentence (Haiman, 1985).

In sound symbolism, we also deal with the high-order similarity where the properties of abstract 'referents' are associated with the elements of a word form, and where the human—cognitive, motor, spatial, emotional, etc. experiences are symbolized by sounds. E.g., "high tones, vowels with high second formants (notably /i/), and high-frequency consonants are associated with high-frequency sounds, small size, sharpness, and rapid movement; low tones, vowels with low second formants (notably /u/), and low-frequency consonants are associated with low-frequency sounds, large size, and heavy, slow movements" (Hinton, Nichols, and Ohala, 1994, p. 10). According to Winter, Perlman, Perry, and Lupyan (2017), words referring to sensory domains (sound, sight, touch, taste, and smell) are more iconic than the words with abstract meanings. The differences between onomatopoeia/sound symbolism are reflected in the two types of iconicity—absolute or primary/relational or secondary correspondingly.

In phonaesthemes, the relations of systematicity are brought to the fore: they possess an initial cluster of phonemes which occurs regularly within a set of words. E.g., the words starting with *bl-, sn-, gl-, pr-*, etc. have some similarity of meaning, referring to such semantic classes as 'audible', 'perceptible', 'moving', etc. (Lvova, 2005). In this case, we are not concerned with a direct relationship between form and meaning but with a systematic correspondence between them. Such "regular mapping between aspects of form and function" (Dingemanse et al., 2015) features distributional regularities in different languages. Various scholars measuring sound/meaning similarity distance (Shillcock, Kirby, McDonald, and Brew, 2001); (Abramova and Fernández, 2016)

have arrived at the conclusion about the ubiquitous character of this phenomenon.

Most linguists study iconic signs within the framework of phonological iconicity. Our aim was to show the role and place of referential meanings in non-arbitrary form-meaning relationships in the history of the Old English language. Indeed, iconicity plays an important role in language development: according to Perniss and Vigliocco (2014), it "bridges between" language and human experiences and "support *referentiality*" (the ability of speakers to label objects and events in the processes of nomination) and "displacement" (the ability of linguistic signs to stand for the referents).

Within the course of language development iconicity erodes: a good example of erosion is the process of grammaticalization where the forms with more concrete meanings are superseded with the forms with more abstract meanings. E.g., the Old English verbs of existence *beon/wesan* used to nominate more concrete meanings of 'growth', 'biding', and 'dwelling' clearly capturing some iconic properties relevant for the speakers. In the course of time these meanings semantically eroded and later were replaced by the more abstract grammatical meaning of 'existence'. One more example of the erosion of iconicity is the language vocabulary: "with vocabulary growth, representational spaces comprising forms and meanings become more densely populated, thereby increasing the possibilities of confusion and ambiguity in the spoken forms of words, providing a selective pressure towards more arbitrary, more discriminable forms" (Dingemanse et al., 2015).

5. Data and Methodology

Our research is performed on the basis of the Old English lexicon. Computational analysis in the field of diachronic linguistics is based on such data preprocessing as lemmatization, stemming, POS and semantic tagging, morphological and syntactic markup. Old English is a low-resource language. The limitation of Old English textual data and the absence of finished implementations for older languages preprocessing constitute considerable challenge. Due to the broad dialectal variation in Old English, there is a large number of orthographic variants, which makes it difficult to perform lemmatization, stemming and POS tagging (the existing Python implementations provide inaccurate results). Morphological and syntactic markup is available only for a small number of texts. Thus, the scope of our study was narrowed, and it was decided to focus on the analysis of A Thesaurus of Old English,[5] with the appli-

5. http://oldenglishthesaurus.arts.gla.ac.uk/

cation of statistical methods of the research. A Thesaurus is provided under the license.

The data comprise the lexemes from A Thesaurus sharing the initial consonantal graphemes (phonemes) <w>, <s>, <h>, and <p>. The lexemes with the initial graphemes <w>, <s>, <h> are the most frequent ones while the lexemes with the initial <p>, on the contrary, are the least frequent. After processing, the dataset for the analysis comprised the lexemes with four onsets: <w> (3359 words), <s> (4586 words), <h> (5210 words), and <p> (542 words) excluding entries with the compound lexemes written separately or with a hyphen (13,697 words in total). In A Thesaurus lexical meanings of Old English words are arranged into 18 conceptually organized semantic categories. The categories are: 1. The Physical World, 2. Life and Death, 3. Matter and Measurement, 4. Material Needs, 5. Existence, 6. Mental Faculties, 7. Opinion, 8. Emotion, 9. Language and Communication, 10. Possession, 11. Action and Utility, 12. Social interaction, 13. Peace and War, 14. Law and Order, 15. Property, 16. Religion, 17. Work, 18. Leisure. A certain semantic category is ascribed to each word, e.g., *wæþ* 'a ford' is given under the category 5. 'Existence'. We hypothesize that there might be some non-random distribution of semantic categories over the lexemes sharing initial graphemes.

Statistical methods are widely used in linguistic research. For the analysis of the distribution the most commonly used one is Pearson's Chi-square, in particular the Chi-square test and the coefficient of contingency. Despite frequent critical remarks on the application of Pearson's Chi-square in linguistics we consider it to be appropriate and reasonable for our dataset though with certain assumptions. Pearson's Chi-square attempts at making a conclusion whether a distribution observed is purely accidental, or whether it reflects a certain regularity. This statistical test is applied to a contingency table made up of the element frequency in the sampling unit to be compared with the total number of elements in this unit. The null hypothesis tested is that the difference between the element frequency is the result of random variations. The implementation of Pearson's Chi-square statistics is available in a number of software frequently used in corpus linguistics such as WordSmith Tools and AntConc but in our case the data preprocessing and Pearson's Chi-square statistics were performed within Python realization.

6. Form-Meaning Hypothesis in a Thesaurus of Old English

For the analysis we calculated the frequency of lexemes starting with <w>, <s>, <h>, and <p> graphemes in every semantic category of A Thesaurus of Old English. The results are presented in Table 1.

TABLE 1. The distribution of the initial graphemes <w>, <s>, <p>, <h> in semantic categories

	<w>	<s>	<h>	<p>
The Physical World	193	335	247	25
Life and Death	553	809	932	109
Matter and Measurement	164	327	261	23
Material Needs	283	478	514	76
Existence	347	707	658	53
Mental Faculties	216	264	222	15
Opinion	146	64	167	26
Emotion	236	202	307	9
Language and Communication	118	157	69	14
Possession	17	34	67	0
Action and Utility	135	150	214	16
Social interaction	304	246	318	14
Peace and War	126	185	234	5
Law and Order	114	106	105	10
Property	75	64	66	11
Religion	228	275	698	75
Work	66	113	88	27
Leisure	38	70	43	34

Table 1 presents a contingency table of two categorical variables—the semantic category and the initial grapheme of the word. To check if the semantic categories are distributed among the lexemes with identical initial graphemes non-randomly, we applied the Chi-square test. The Chi-square statistics is commonly used for testing relationships between categorical variables. The Chi-square test measures the ratio of difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table, which enables to estimate the relationship between the variables. The null hypothesis of the Chi-square test is that the categorical variables are statistically independent. The null hypothesis will be recognized when the observed frequencies are less than the expected counts. In cases where the observed frequencies are greater than theoretically expected ones, the relationships between variables are statistically significant and present evidence for correspondence between them. The formula for the Chi-square test is:

$$\chi_c^2 = \sum_i \frac{(O_i - E_i)^2}{E_i},$$

where c = degrees of freedom, O = observed value(s), E = expected value(s).

The application of the Chi-square test to the resulting table shows the following statistics: $\chi^2 = 765.67$, $p = 3.637 \times 10^{-128}$, $df = 51$. The critical value of the Chi-square statistics with $50 < df < 55$ varies from 67.51 up to 86.66 and more. Thus, the resulting statistics can be considered an example of non-random variation with the significant p-value. Thereby, we can conclude that the null hypothesis of the independence of variables can be rejected.

Statistics also offers to analyze the cell-wise contributions to the Chi-square to see where the evidence for the dependence is coming from. The contribution to the Chi-square quantifies the individual category contributions, i.e., how much of the total the Chi-square statistic is attributable to each category difference between observed and expected values. The contribution to the Chi-square is found by taking the squared difference between the observed count and the expected count then dividing by the expected count. The results of the contributions to the Chi-square are presented in Table 2. Larger values indicate a more substantial contribution to the overall Chi-square statistics.

Table 2. Contributions to the Chi Square for Old English initial <w>, <s>, <p>, <h> in 18 semantic categories

	<w>	<s>	<h>	<p>
The Physical World	0.05	16.84	10.79	1.41
Life and Death	2.23	0.02	0.35	2.01
Matter and Measurement	3.57	17.57	3.87	1.93
Material Needs	7.04	1.46	0.00	9.46
Existence	17.02	22.80	0.26	4.09
Mental Faculties	9.18	2.39	9.43	6.32
Opinion	22.52	37.28	1.23	6.31
Emotion	14.12	10.08	1.42	14.58
Language and Communication	10.39	11.51	33.13	0.00
Possession	5.00	0.81	10.67	3.75
Action and Utility	0.60	2.92	1.68	0.95
Social interaction	35.57	8.23	0.91	12.54
Peace and War	0.58	0.00	2.94	12.93
Law and Order	12.35	0.34	3.95	0.80
Property	9.16	0.96	3.18	0.70
Religion	23.04	54.23	93.18	11.84
Work	0.52	2.16	5.08	20.25
Leisure	1.20	1.05	10.64	97.10

To estimate the relationship of every initial grapheme to certain semantic category we computed the Chi-square test for alternative dis-

tribution contingency 2 × 2 tables compiled on the basis of Table 1 as follows:

TABLE 3. The alternative distribution contingency table

	religion	other semantic categories	total
<p>	75	467	542
other initials	1,201	11,954	13,155
total	1,276	12,421	13,697

In cases where the association between two variables can be statistically significant (the Chi-square and p-value results), the strength of this association can be very small. To decide whether the relationship between two categorical variables is important, the coefficient of contingency (Phi (or φ)) is computed. The Phi coefficient measures how closely the observations in rows and columns are associated with each other. The formula for the Phi coefficient is $\varphi = \sqrt{\frac{\chi^2}{n}}$, where n is the total number of observations.

Table 4 provides the Chi-square, p-value and Phi coefficient results only for the cases where the observed data exceed the expected figures. The authors reason such data elimination by focusing on positive correlation between the variables. The critical value of the Chi-square statistics with $df = 1$ is 3,84. The range of Phi values from 0.05 to 0.1 are considered to estimate moderate positive relationships.

The results of the Chi-square together with the coefficient of contingency and the contributions to the Chi-square for the words with the initial graphemes <w>, <s>, <p>, <h> turn out to be different for certain semantic categories. This testifies to the idea of the correspondence between the initial grapheme of the word and its semantics. The resulting values allow the authors to reject the null hypothesis as they manifest non-arbitrary association between the words starting with different onsets and semantic categories:

<w> 'Social interaction' (*wealdan* 'to rule', *werod* 'assembly'), 'Opinion' (*wlanc* 'arrogant', *wlitig* 'fair, noble'), 'Emotion' (*weorc* 'suffering', *wynsum* 'joyful'), 'Law and Order', 'Language and Communication', 'Mental Faculties', 'Property'.

<s> 'Existence' (*sālnes* 'a time of silence', *samod* 'without a break'), 'Matter and Measurement' (*samen* 'together', *sid* 'large'), 'The Physical World' (*seolfor* 'silver', *sicel* 'small stream').

<h> 'Religion' (*halig* 'holy', *hals* 'salvation'), 'Possession' (*habban* 'to possess'), 'Peace and War.

TABLE 4. Chi Square p-value and Phi coefficient data for Old English initials <w>, <s>, <p>, <h> in 18 semantic categories

	<w> chi^2, p-value, ϕ	<s> chi^2, p-value, ϕ	<h> chi^2, p-value, ϕ	<p> chi^2, p-value, ϕ
The Physical World		26.87 2.17E-7 0.04		
Life and Death		0.045 0.83 0.002	0.69 0.4 0.007	2.57 0.1 0.01
Matter and Measurement		27.99 1.22E-7 0.05		
Material Needs		2.43 0.12 0.01	4.45E-5 0.99 5.70E-5	10.98 0.0009 0.03
Existence		39.33 3.59E-10 0.05		
Mental Faculties	12.83 0.0003 0.03	3.79 0.052 0.02		
Opinion	30.73 2.96E-8 0.05 / 5%		2.04 0.15 0.01	6.8 0.009 0.02
Emotion	19.79 8.63E-6 0.04		2.43 0.12 0.01	
Language and Communication	14.14 0.0001 0.03	17.76 2.51E-5 0.04		
Possession			17.74 2.53E-5 0.04	
Action and Utility	0.83 0.36 0.008		2.8 0.09 0.01	
Social interaction	50.36 1.28E-12 0.06			
Peace and War		0.006 0.94 0.0007	4.94 0.03 0.02	
Law and Order	16.76 4.23E-5 0.03			
Property	12.33 0.0004 0.03			0.74 0.39 0.007
Religion			165.79 6.13E-38 0.11	13.66 0.0002 0.03
Work		3.31 0.069 0.02		21.59 3.37E-6 0.04
Leisure		1.6 0.2 0.01		102.62 4.05E-24 0.08

<p> 'Leisure'(*pipe* 'a flute', *plega* 'play', 'sport'), 'Work' (*pāl* 'a spade', *pinn* 'nail'), 'Religion' (*preost* 'a priest', *postol* 'an apostle'), 'Material Needs', 'Opinion'.

7. Conclusion

In our study we have made an attempt to test the form-meaning hypothesis in earlier stages of language. The outline of our research sketches the entire lexicon, not the localized phonaesthemic patterns which are mostly examined in the related studies. To explore such patterns the authors apply the methods of historical semantics for thorough linguistic analysis without reference to computational tools.

We analyzed the methods of computational linguistics and realized that in relation to older languages not all of them are suitable. We have decided to employ statistical methods for our research. Although there is a lot of criticism about the use of statistical methods in linguistics, the results are worth considering as the application of these methods is promising for identifying the regularities in language development. We computed the relationship of Old English words sharing consonantal initial graphemes (phonemes) and their conceptual (referential) meanings in the lexicon. We found out that there is a certain association between an initial grapheme and semantic category to which the word sharing this grapheme belongs. We have revealed a regularity that the distribution of semantic categories among the words starting with one initial grapheme differs from the distribution of semantic categories among the words starting with the other onsets. We assume that this regularity may take place by chance, but it is highly likely to be based on non-arbitrary form-meaning relations. This patterning may be random, but it may as well be determined by the iconic coding in language.

For a while, the form-meaning hypothesis was tested only on four initial graphemes in the Old English lexicon. We expect that for a larger number of graphemes the results would be different. In the future, we plan to expand the dataset with more initial graphemes and to undertake a full-scale research. Further explorations into the topic can be continued in the sphere of semiographemics since form-meaning relationships can be also traced in other semiotic modelling systems (symbolic writings and art) with specific implications for form-meaning hypothesis (Lotman, 2011).

References

Abramova, Ekaterina and Raquel Fernández (2016). "Questioning Arbitrariness in Language: a Data-Driven Study of Conventional Iconicity." In: *Proceedings of the NAACL-HLT, San Diego, California*, pp. 343–352.

Altmann, Gabriel (2004). "Script Complexity." In: *Glottometrics* 8, pp. 68–74.

Aryani, Arash, Arthur M. Jacobs, and Markus Conrad (2013). "Extracting Salient Sublexical Units from Written Texts: "Emophon," a Corpus-Based Approach to Phonological Iconicity." In: *Frontiers in Psychology* 4.654.

Atkinson, Mark, Gregory J. Mills, and Kenny Smith (2019). "Social Group Effects on the Emergence of Communicative Conventions and Language Complexity." In: *Journal of Language Evolution* 4.1, pp. 1–18.

Barney, Stephen A. (1985). *Word-Hoard: An Introduction to Old English Vocabulary (Yale Language Series)*. 2nd ed. Yale: Yale University Press.

Blasi, Damián E. et al. (2016). "Sound-Meaning Association Biases Evidenced across Thousands of Languages." In: *Proceedings of the National Academy of Sciences*. Vol. 113, pp. 10818–10823.

Cornell, Muriel (1981). "Varieties of Repetition in Old English Poetry. Especially in The Wanderer and The Seafarer." In: *Neophilologus* 65.2, pp. 292–307.

Dingemanse, Mark et al. (2015). "Arbitrariness, Iconicity, and Systematicity in Language." In: *Trends in Cognitive Sciences* 19.10, pp. 603–615.

Fulk, Robert D. (2014). *An Introductory Grammar of Old English with an Anthology of Readings*. 1st ed. Vol. 463. Medieval and Renaissance Texts and Studies. Tempe, Arizona: ACMRS Press.

Givón, Talmy (1985). "Iconicity, Isomorphism, and Non-Arbitrary Coding in Syntax." In: *Iconicity in Syntax. Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24–26, 1983*. Ed. by John Haiman. Amsterdam: Benjamins, pp. 187–219.

Godden, Malcolm R. (1992). "Literary Language." In: *The Cambridge History of the English Language. Vol. I: The Beginnings to 1066*. Ed. by Richard M. Hogg. Cambridge: Cambridge University Press, pp. 490–535.

Gutiérrez, E.Darío, Roger Levy, and Benjamin K. Bergen (2016). "Finding Non-Arbitrary Form-Meaning Systematicity Using String-Metric Learning for Kernel Regression." In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Vol. 1. URL: https://www.aclweb.org/anthology/P16--1225.pdf.

Haiman, John (1985). *Natural Syntax: Iconicity and Erosion*. Cambridge: Cambridge University Press.

Hinton, Leanne, Johanna Nichols, and John J. Ohala, eds. (1994). *Sound Symbolism*. Cambridge University Press.

Jee, Hana, Monica Tamariz, and Richard Shillcock (2018). "The Substructure of Phonics: The Visual Form of Letters and their Paradig-

matic English Pronunciation are Systematically Related." PsyArXiv https://psyarxiv.com/n85mb/.
Jespersen, Otto (1922). *Language: Its Nature, Development and Origin*. London: George Allen & Unwin.
Kastovsky, Dieter (1992). "Semantics and Vocabulary." In: *The Cambridge History of the English Language. Vol. I: The Beginnings to 1066*. Ed. by Richard M. Hogg. Cambridge: Cambridge University Press, pp. 290–408.
Lotman, Yuri (2011). "The Place of Art among Other Modelling Systems." In: *Sign Systems Studies* 39.2/4, pp. 249–270.
Lvova, Nadija L. (2005). "Semantic Functions of English Initial Consonant Clusters." In: *Glottometrics* 9, pp. 21–28.
Minkova, Donka (2003). *Alliteration and Sound Change in Early English*. Cambridge: Cambridge University Press.
Monaghan, Padraic et al. (2014). "How Arbitrary is Language?" In: *Philosophical Transactions of the Royal Society B—Biological Sciences* 369.
Perniss, Pamela and Gabriella Vigliocco (2014). "The Bridge of Iconicity: from a World of Experience to the Experience of Language." In: *Philosophical Transactions of the Royal Society B* 369.
Philps, Dennis (2008). "Submorphemic Iconicity in the Lexicon: a Diachronic Approach to English 'gn-words'." In: *Lexis* 2. URL: http://journals.openedition.org/lexis/728.
Pimentel, Tiago et al. (2019). "Meaning to Form: Measuring Systematicity as Information." In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1751–1764.
Pollington, Stephen (1993). *Wordcraft: New English to Old English Dictionary and Thesaurus*. 2nd ed. Hereward, UK: Anglo-Saxon Books.
Shillcock, Richard et al. (2001). "Filled Pauses and their Status in the Mental Lexicon." In: *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*, pp. 53–56.
Slaměníková, Tereza (2019). "On the Nature of Unmotivated Components in Modern Chinese Characters." In: *Proceedings of Graphemics in the 21st Century, Brest 2018. Ed. by Yannis Haralambous*. Ed. by Yannis Haralambous. Brest: Fluxus Editions, pp. 209–226.
Smith, Chris A. (2016). "Tracking Semantic Change in fl-monomorphemes in the Oxford English Dictionary." In: *Journal of Historical Linguistics* 6.
Smith, Jeremy J. (2009). *Old English: A Linguistic Introduction*. 1st ed. Cambridge: Cambridge University Press.
Tsur, Reuven (2002). "Aspects of Cognitive Poetics." In: *Cognitive Stylistics—Language and Cognition in Text Analysis*. Ed. by Elena Semino and Jonathan Culpeper. Amsterdam, Philadelphia: John Benjamins, pp. 279–318.
Wilcox, Sherman (2004). "Cognitive Iconicity: Conceptual Spaces, Meaning, and Gesture in Signed Language." In: *Cognitive Linguistics* 15.2, pp. 119–147.

Winter, Bodo et al. (2017). "Which Words are Most Iconic? Iconicity in English Sensory Words." In: *Interaction Studies* 18.3, pp. 430–451.

Youn, Hyejin et al. (2016). "On the Universal Structure of Human Lexical Semantics." In: *Proceedings of the National Academy of Sciences of the United States of America* 113.7, pp. 1766–1771.

Zlatev, Jordan, Przemysław Żywiczyński, and Sławomir Wacewicz (2020). "Pantomime as the Original Human-Specific Communicative System." In: *Journal of Language Evolution* 5.2, pp. 156–174.