WUT 2020
10th International Conference "Word, Utterance, Text: Cognitive, Pragmatic and Cultural Aspects"

RELATIONSHIP BETWEEN ENTROPY AND INFORMATIVITY IN THE PROCESSES OF DERIVATION

Balakin Sergey Vladimirovich (a)*
*Corresponding author

(a) Ural State University of Railway Transport, Russia, Ekaterinburg, Syromolotova St., 25 - 12,
sergiorusso2007@mail.ru

Abstract

The modern derivative theory is a synthesis of traditional understanding of the problem and actively developing cognitive approach in linguistics, which forms a methodological basis for the study of linguistic derivation. In this case, the interaction of language structures during changes is being discussed in terms of the formation of a new element in the structure, and during the formation of a new derivative increases the information component of the new word itself. According to this, realization of the derivative and interpretive potential consists in increasing the information component of the secondary derivative lexical unit in comparison with its derivative basis. As studies show, from the point of view of the information component, the primary unit has greater uncertainty, which is reflected in increased entropy. Secondary units formed synthetically rather than semantically have less entropy. The study also showed that entropy directly affects the informativity of both the derivative and the discourse in which this derivative is used. At the same time, with a slight increase in entropy, the information coefficient decreases by several times. The prospects of studying the derivative potential of language lexical systems are seen in our studies of the development of conceptual areas both in isolation and in comparison of different linguocultures. The developed methodology of derivative potential determination can provide an impetus to research of derivative processes at conceptual and linguistic levels.

Keywords: Discourse, derivational processes, semantic development, entropy, informativity, pseudo-information.

1. Introduction

Any text, both official and private, as well as any fictional, journalistic and other texts have a certain degree of informativity. Along with the literary component, the reader should receive not only aesthetic pleasure from reading, but also get certain knowledge. Thus, drafting any text is a complex and multifaceted process that requires some effort from its author. One aspect that can be difficult for any author is the desire to be more precise and specific. That is, in this case we are talking about the fact that the information provided to a text recipient should not be ambiguous, contradictory.

2. Problem Statement

The problem of defining informativity is important. This applies especially to official papers (if they can be classified as a work of language). A lot depends on how a document is drafted: its understanding, correctness of actions, etc. Misunderstanding of the text can lead to negative consequences. It is considered that the problem of determining informativity is not only a linguistic but also a mathematical problem. The informativity variation of any text (written and oral) leads us to two problems. The first one is linguistic one, which covers the range of problems associated with the linguistic embodiment of any text and, accordingly, it is necessary to determine the factors that influence informativity. The second one is mathematical, which is trying to decide how much of the informativity of a text suffers if it (text) is compiled incorrectly. The second problem seems to be related to the notion of entropy. These both problems are inseparable and their connection is formed at the syntactic level. If we imagine that the syntactic structure of the text has a propositional basis, then we should talk about the formation of informativity at the level of propositional relationships within each proposition. Thus, it is necessary to determine the following fact: how relevant the proposition is when the language is striving for complete informativity, on the one hand, and linguistic economy, on the other. Entropy is defined as the amount of information per language symbol, expressed both in spoken and written forms. The amount of information is calculated as a product of entropy by the number of symbols written (spoken) in the text: the bigger this number, the more information and, accordingly, the more entropy. A decrease in the entropy coefficient indicates an increase in the uncertainty of the sentence.

3. Research Questions

According to Claude Shannon, entropy is a statistical parameter, which is measured in a certain sense by the average amount of information per letter of a language text (Shannon, 1963). The scientist used entropy in the experiment: a possibility to predict the English text, how precise the prediction of the English text can be when the previous N letters of the text are known (Shannon, 1963). For example, the phrase *someone is going to come out of the doors of the building* is uninformative and has low entropy, as we do not know who is going to come out of the doors of the building (man, woman, child, etc.). Informativity increases when you specify what kind of building it is. For example, «in the phrase *someone is going to come out of the doors of the **barracks** building* it is most likely to be a male» (Shannon, 1963, p. 216). Informativity is calculated using the following Shannon's formula: $I = -(p_1\log_2 p_1 + p_2\log_2 p_2 + \ldots + p_N\log_2 p_N)$, where p_i is the probability that the *i*-th message is selected in a set

of N possible derivatives. Thus, the formula is given as follows: $N = -\sum_{i=1}^{N} p \log p$, where N *is* the coefficient of possible variations of the primary and secondary units; $-\sum_{i=1}^{N} p \log p$ is the sum of all possible variations in the derivative (for example, walk - go - cross, etc.).

Thus, entropy is directly related to the text informativity (with its meaning) in which a certain derivative is used. It is considered that if entropy decreases, the information content of the text also decreases. However, if one part of the information in the text is ambiguous, then informativity fills this gap in the other part of the text (Balakin, 2018).

Balakin (2015a) explored the entropy of the propositional structure is possible on the word-formation level, for example, by varying different prefixes. For example, есть – доесть – переесть – съесть, etc. So entropy is stronger if there is already a prefix that carries certain information about the derivative. Conversely, if there is no prefix, there is less entropy because it is possible to use a derivative with another prefix. In this case, entropy can be increased on a discursive level, as language units surrounding the derivatives can make up for the lack of information (Balakin, 2015a).

In the French texts listed below prefixal derivatives perform the function of packaging the following information. 1. Time (repetition), 2. Reverse action, 3. Confined space, deprivation, weakening, 4. Result. There are some examples of the prefixal reflection of information in Portuguese discourses: 1. Confined space, opposition, 2. Negation, 3. Time (antecedence), 4. Deprivation, 5. Division, 6. Opposition (Balakin, 2015c; Balakin, 2015d; Balakin, 2015f; Balakin, 2015h).

When the folded elements are more than necessary for understanding, there may be a crisis point of entropy reduction that can be taken away at the expense of context as a result of the high concentration of additional elements in the conceptual structure. Such concentration, according to Nefedova (2013), is «achieved due to the implicitness that allows redistributing the cognitive load in particular conceptual structure, and is motivated by the selectivity of the language to the signs of extra-linguistic reality». So, it is necessary to find out not only the types of proposition foldings, but it is also important to predict what is the probability of entropy reduction and how it will affect the text informativity, how the text behaves in case of low entropy. From the derivative processes point of view, entropy is folding of the proposition with a positive sign. The derivative itself is reduced to one word, i.e. it can be said that it is a foldability (regression) is also with a positive sign. Entropy reduction (i.e. an increase in the number of variations or omission of word-formation elements) leads to ambiguity of the information contained in the derivative, which may result in a destructor of the meaning of differentiation and understanding. Thus, we consider entropy as a variation of elements in proposition in derivative processes, in other words, which one element (characteristic) is used in the structure and by which language unit it is verbalized (Balakin, 2015a; Balakin, 2015b, 2016d).

4. Purpose of the Study

To determine the derivative potential general patterns of the lexical language systems realization on the basis of the study of relations between the primary and secondary units.

To determine natural internal properties of propositional structure in derivative processes. To develop the principles of mathematical modeling of the mutual influence of entropy and informativity of

secondary language units on the basis of the propositional method (Balakin, 2015a; Balakin & Nefedova, 2016c; Balakin & Ankov, 2015g; Balakin, 2016a; Kravchenko, 2015).

5. Research Methods

The fundamental principle of the work is the conceptual method in the analysis of ontological characteristics involved in derivative processes (Balakin, 2016b; Dicad de Português, 2015), the propositional method, the mathematical-statistical method (in the study of entropy and the informative component in derivative processes).

6. Findings

The mathematical analysis of calculation of informativity and entropy shows that when the position is critically folded, the first thing to be lost is the semantic part of the text (its informativity) containing derivatives. It is estimated that when the prefix in the derivative is intentionally removed, entropy reduces. At the same time, the more derivatives with remote prefixes there are, the less entropy there is. Let's compare the texts *La mort heureuse* (Camus, 2010) and *A Bruxa de Portobello* (Coelho, 2006).

For example, Balakin (2018) says in the following texts, in the artificial omission of the conceptual characteristics of "reverseness", "phase", "locativity", expressed by means of prefixes dé-, en-, ex-, o-, про-, пере-, respectively, the phrase entered the phase of reduced entropy and pseudo-information.

Balakin (2018) explored the previous text in the initial (not transformed) version has the text informativity coefficient 552 whereas after transformation this coefficient is 530 which is 0.96 less that in the initial one. The entropy of untransformed and transformed texts is 4 and 3,8 respectively, which is 0.98 difference. Therefore, we can conclude that entropy reduces the informativity of the text by about 2 points (that is 0.98 - 0.96 = 0.02).

In the example given the coefficient of informativity of the nontransformed and transformed text makes 271 and 260 respectively, that is the coefficient of informativity loss makes 0,94. Meanwhile, the entropy coefficient in the text under study is 4 and 3. The difference in the coefficients is 0.92. This difference reflects the fact of information loss during regressive fold by one (1) element. To prove it, let us give an example with a similar prefix, but the length of the analyzed text will be longer (Balakin, 2018).

Balakin (2018) explored the entropy change coefficient in the given text is 0.96, and the informativity change coefficient is 0.94.

At calculation of average quantity of four examples it was found out that reduction in entropy occurs on 0,02, and average decrease in informativity of the text thus makes 0,05.

7. Conclusion

The mathematical method of calculating the influence coefficients of the derivative entropy on the discourse informativity showed that if the entropy reduces by 1%, the informativity on average decreases

by 2.5%. Thus, informativity turns out to be most dependent on the entropy of a derivative, i.e. derivative processes prove to be very important to follow from, on the one hand, language economy and, on the other hand, semantic completeness of the secondary unit point of view. As a result of reduced entropy and informativity, pseudo-information and discourse implicitness may arise, that can be overcome by including more context. This confirms another property of the proposition: its homogeneity. Homogeneity also enables it to be implemented in different linguistic morphological and syntactic units simultaneously. Cognitive homogeneity in derivation is clearly visible in the process of diachronic language analysis.

According to this principle we distinguish three types of folded propositions: 1) folding up to one element; 1) folding up to two elements; 2) folding up to three elements. Within each type there is also a fold gradation that depends on the derivation kind. In suffixing and word-complexation, the folded proposition is of the first type, but the predicate is more detailed when forming complex derivatives or composites. The second type is present at derivations of prefixal-suffixal and prefixal types, with the prefixal type showing greater foldability. Specific features, however, depend primarily on the very structure of the target languages. Thus, first of all, the Russian language has a greater folding potential due to the profiling of derivational nodes in the affix fund; Portuguese and French tend to use lexico-syntactic constructionse (Balakin, 2015e).

References

Balakin, S. V. (2015a). Mathematical and linguistic modeling of the synthetic derivation conceptual bases. *USURT Vestnik*, *2*(26), 57–70.

Balakin, S. V. (2015b). Derivation Processes of Detailed Elaboration and Intensification in The Structure Macro Concepts. *Cognitive language investigations. Issue XXII. Language and consciousness in the interdisciplinary research paradigm*, 193–195.

Balakin, S. V. (2015c). Fundamentals and principles of derivation at the conceptual level. *Seminar of doctoral candidates of USURT: a collection of scientific papers*, 5–21.

Balakin, S. V. (2015d). Derivational Potential of Predicates of The First Sequence (Based On The Predicates "être", "avoir", "aller", "mettre", "faire" In French). *Vestnik of university (Russian-Tajik (Slavic) University). Scientific journal*, *2*(49), 204–212.]

Balakin, S. V. (2015e). Propositional Basis Of Formation Of Prefixed Verbs In French And Portuguese. *Scientific notes of the Orlov State University*, *6*(69), 92–96.

Balakin, S. V. (2015f). Formation of verbs with prefixes from a substantive basis in Portuguese language (cognitive aspect). *Vestnik of Nizhni Novgorod State Linguistic University named after N.A. Dobrolyubov*, *29*, 9–19.

Balakin, S. V. (2016a). Principes dérivatives et cognitives de la sphère verbale russe. *Cadernos de Linguagem e Sociedade*, *17*(2), 154–168.

Balakin, S. V. (2016b). Study of the derivative processes in the context of the conceptual-taxonomical method. *Scientific review: Humanitarian research*, *10*, 206–218.]

Balakin, S. V. (2018). *Derivation potential of lexical systems of language*. (Thesis).

Balakin, S. V., & Ankov, A. A. (2015g). Nominative And Derivation Potential Of Concept "Invective" (By The Example Of The French Linguistic-Culture). *Philological sciences. The questions of theory and practice*, *3*(45), 24–27.

Balakin, S. V., & Nefedova, L. A. (2016c). Perspectives of conceptual research in derivatology. *Concept and culture: a collection of articles (VI International Scientific Conference, Kemerovo - Yalta, 25-27 September 2016)*, 78–83.

Balakin, S. V. (2015h). Conceptualization at the analytical language derivation. *Cognitive language studies. Issue XX. From cognitive linguistics to cognitive terminology*, 758–766.]

Balakin, S. V. (2016d). Basic postulates of a cognitive derivatology. *Seminar of doctoral candidates of USURT : a collection of scientific papers*, 5–11.]

Camus, A. (2010). *La mort heureuse* [Happy death]. Gallimard.

Coelho, P. (2006). A Bruxa de Portobello [The Witch of Portobello]. Editora Planeta.

Dicad de Português (2015). Morphology and Phonology. National Council of Justice.

Kravchenko, S. Yu. (2015). Prototypical Approach and Process Of Categorization. *Philological sciences. The questions of theory and practice*, *3*(45), 117–119.]

Nefedova, L. A. (2013). Cognitive characteristics of the implicit presentation of art. *Bulletin of Chelyabinsk State University. Philology*, *24*(315), 31–34.

Shannon, C. (1963). *Information theory and cybernetics*. Foreign Literature Publishing House.