Linguistics and Cryptophilology*

Anton Zimmerling

Pushkin State Russian Language Institute / Institute of Linguistics
Russian Academy of Science
fagraey64@hotmail.com

Abstract. I compare language and text as research objects and discuss the status of text linguistics aka cryptophilology. The latter exploits the philological concept of text, which is masked by the use of linguistic terms. Linguistics deals with automatically realized language structures, while cryptophilology explains how a unique speech situation is displayed in the composition of a unique text.

Keywords: linguistics, philology, text, levels of representation, units of segmentation.

1. Identifying science(s)

The branches of science can be identified in a threefold way: 1) restrictively, by providing a narrow definition of their research object – e.g., if S is a branch of science, its restrictive definition must tell what S is *not* about; 2) non-restrictively, by a bunch of specific *methods* characteristic of S; 3) as a kind of activity driven by people who *call themselves S-scientists*. A pre-theoretic intuition or public reaction to the existence of people doing linguistic research professionally is that linguistics is a science branch L, which deals with natural language phenomena. The notion of natural language(s) is at first glance simple. It can either be reduced to presumably more basic notions like 'communication' and defined, e.g. as 'primary means of communication in a community of speakers' or taken to be elementary, i.e. not requiring any definition. A pre-theoretic intuition concerning natural language is that it is not the same thing as, e.g. art, cognition, or behavior. However, the methods implemented in the description of languages like English, French, Latin, Warlpiri, etc. during the last 70-80 years have been applied to art, cognition, behavior, and other objects presumably related to but not identical with natural language. This makes a method-oriented approach to defining L confusing. Either L lacks methods specific for it or exports them to some art and social sciences. The diversity of

---

schools claiming the names of *corpus linguistics*, *psycholinguistics*, *poetic linguistics*, *theological linguistics*, *political linguistics*, *forensic linguistics*, etc. is puzzling.[1] Let us assume that every expression of the type *l-linguistics* refers to a domain $L_k \in \{L_1, L_2...L_n\}$. The perspective of reconstructing the unity of L by a set-union operation $L_1 \cup... \ L_n$ is obscure, since the elements from L can represent different sciences, not a single one. A brief look at the methods implemented by some highly specialized kinds of l-linguistics, e.g. calculating the Pearson correlation ratio, fixing the eye movement in a dialogue interaction, or estimating the time distance from a common ancestor by making phylogenetic trees, seems to confirm the pessimistic insight that l-linguists borrow their methods from mathematical or natural sciences. But even if optimists like Jakobson (1973) are correct and L indeed lends its methods to some art and social sciences, we still need a restrictive definition of L.[2] The crucial point is the boundary between language and *text*, and the sciences dealing with them – L and *philology*, respectively. In the following, I adopt the standard philological concept of text as fixed verbal form. This traditional use deviates from the use of the term *text* in the MTT tradition (Mel'čuk 2016), while the use of the term *language* is hopefully consistent with it.

2. Language and text

L divorced philology only in the XXth century when Saussure and his followers established the inventory of language units and proposed a multi-level representation of language. Benveniste (1962) argues there are no language units above sentences, since a sentence as a text segment does not make up a class of sentences and cannot be made an element of any structure of a higher level. I believe this is correct if one stipulates that complex sentences and other poly-predicate complexes have a syntactic structure of their own. The upper limit of the linguistic segmentation is arguably a more basic condition than multi-level theories of grammar, since the same condition holds for modular theories, where well-formed syntactic structures are derived by parallel operation of several modules. Contrary to mainstream linguists, who assemble text fragments from below, i.e. proceed from the minimal language units up to the maximal ones, mainstream philologists do the segmentation from above: they treat the entire text as a unique object and follow its structure down to the limit, where they can assign a compositional value to its fragments. It is hardly surprising that such minimal unit of a text segmentation from above has been introduced as *texteme*, an innovative term echoing the X-*eme* terminology in linguistics, cf. *phoneme*, *morpheme*, *toneme*, *phraseme*, etc.[3] The name of the dissertation Сибиряков (2020) referring to "texteme organization of a text in a certain group of texts" may look bizarre, but the motivation of this approach is generally clear – it attempts at finding the text units characteristic of some author,

---

[1]   Neither of these terminological combinations is coined by the author of this paper: there is a bulk of professional research behind each of them.

[2]   An opportunistic wisdom that is seldom proclaimed openly but remains popular on the sidelines is that L is what X (the speaker and / or his school) considers relevant for L. If one takes this argument seriously, it just means that X (= an l-linguist) knows a restrictive definition of L.

[3]   The term *dicteme* proposed in Блох (2000) seems to be a variant to the more traditional rhetoric terms like *superphrasal unity*, *period of speech*, *paragraph*. This is borderline notion between L and philology. On the one hand, Bloch (*ibid.*: 58) claims that that dictemes have syntactic organization. On the other hand, the text segmentation into dictemes / paragraphs comes from above, not from below.

some genre or some mentality shared by those who apply to a certain text genre.

The depth of the philological analysis is amazing – in some cases, the philologists can explain the choice of an allomorph by the special intention of the author, who lived several hundred[4] or even thousands of years ago – which gives an overt impression that L and philology are different facets of the same science. This is an illusion. The main problem is not that philologists should get rid of linguistic skills and forget linguistic terms like *allomorph*, but that in many cases a linguist must put aside the philological concept of text. The antagonism of L vs philology is a product of a recent time. Earlier, only philologically-informed linguists capable of retrieving historical and philological information from the text were regarded as valuable members of the science community. Nowadays, philologically-informed linguists constitute a minority, and the call to separate apples and oranges may look weird. Nevertheless, I will discuss the perspective of eliminating the philological concept of text from L. Philologists describe texts as unique objects and provide them with type characteristics in terms of genre systems, culture-specific concepts, etc. Linguists deal with regular mechanisms realized automatically due to some underlying principles of language structure or to presumably universal systemic relations that hold for all world's languages and all authors. L explains predicative structures and their elements and does not dig into text analysis.

3. Cryptophilology

There is a bridge between L and philology, known as *text linguistics* (TL): it claims the right to analyze syntactically amorphous text fragments that do not represent any paradigmatic classes of units. In the last quarter of the XXth century, many authors welcomed the recognition of a coherent text as a linguistic object – cf. Daneš (1974), Падучева (1996: 195), Блох (2000: 61). However, I argue this is a burden: if one gives up the upper limit of segmentation and digs in the context ad infinitum, the border line between regular language meanings and meanings representing the free choice of the speaker easily gets lost. According to the criterion adopted here, TL is classified with philology, but its self-designation hints that people active in this field are prone to classify it with L. The motivation for this decision comes from the fact that TL can be formalized up to the same extent as L, but only one group of philologists is interested in doing that. Two different groups of philologists can be called traditional scholars and cryptophilologists. The former focus on hermeneutic issues and do not apply to the notions of modern L. The latter, cf. such famous names as Andreas Heusler (1923, 1969), Mikhail Steblin-Kamenskij (2003), Olga Smirnickaja (1994), Elena Paducheva (1996, 2018), adapt part of the linguistic apparatus and develop their accounts of *historical linguistic poetics*, *narratology*, *oral literature* involving the application of both philological and linguistic analyses. Henceforth, I refer to TL as cryptophilology.[5] This branch is not devoted to literary traditions and fiction texts but includes a great deal of modern pragmatics explaining the build-up and progress of a text, cf. the speech act theory (Austin 1962; Searle 1983) and dialogue models (Grice 1975), cf. also Sidorov et al. (2014). The classification of speech acts and intentions echoes the classification of text genres. The theory of

---

4   Cf. Gippius (1996: 50) for the choice of the variant of the Gen.Sg.F inflexus -ᴀ /-ja/ instead of -ѣ /-je ~ -ji/ in the XIIIth century Novgorod.

5   Cryptophilogists often retain an active interest to L, which was the case with all prominent scholars mentioned above.

topic-focus articulation (Mathesius 1939) is a bi-product of speech act theory, since different types of topics and foci are diagnosed based on the question-response pairs (Bally 1932; Янко 2008). Moreover, communicative status, as well as such linguistic parameters like word order, intonation, use of applicative morphemes, particles, and special syntactic cues like cleft, pseudo-cleft, and presentational constructions, link sentences with discourse contexts (Ковтунова 1976; Lambrecht 1994). It is difficult to classify all studies of communicative structure with L, since part of the researchers project the notions of topic and focus to text level and assign the corresponding tags to large text fragments lacking a clear-cut syntactic structure, cf. Daneš (1974). In the next section, I specify that the notion of context still can be a legal part of the linguistic procedure, but this requires a carefully defined research task.

4. Context

A problem with the restrictive approach to L is that the end distribution of morphological forms is sometimes established after the look-up of large text fragments (Paducheva 1996: 158). This is characteristic of the so-called differential argument marking (Aissen 2003) when the assignment of morphological case depends on such factors as referential properties of the nominal expression or its communicative status.[6] The inverse connection Pragmatics → Grammar justifies the recourse to the notion of the context in L. Another complication is that some meanings are rather linked with language structures or with operations changing language structure than directly with language units. E.g., Slovenian clustering auxiliaries and pronominal clitics normally take clausal-second position and behave as enclitics, cf. $Videl_{\text{SEE.PERF.SG}} = sem_{\text{BE.AUX.PRES.1SG}} = ga_{\text{ACC.SG.M/N}}$ 'I have seen him / it', but can be fronted in emphatic contexts, where the same elements behave as proclitics: $\#Sem = ga = videl$ 'I have indeed seen it'. The CL1 orders in Slovenian are context-bound and, in addition, register-bound: they occur in oral discourse. I am nevertheless inclined to acknowledge emphasis or *theticity*, i.e. special communicative status that does not bring about a bi-partition of the sentence into topic and focus (Sasse 1987) a regular language meaning and not optional text characteristics. The problem is that theticity is not an inherent feature of any lexical or syntactic item (verbs, clitics, pronouns, auxiliaries). This feature is only activated if the word order shift V — CL ⇒ #CL — V takes place. In Slovenian, the clitic fronting CL2 ⇒ CL1 is a means of converting the basic topic-focus structure into a thetic structure. Similar operations are attested in the class of the world's languages sharing some morphosyntactic conditions with Slovenian (Zimmerling 2021: 527). Communicative meanings like emphasis, topicality, theticity, contrast, etc. can only be diagnosed in contexts that show the distribution of language structures expressing these meanings in a world's language. I conclude that a look-up of the context is a legal part of the linguistic procedure, but the linguistic notion of context (≈ set of necessary or sufficient conditions for the realization of regular language meanings) is different from the philological concept of text as fixed verbal form. If one however treats the situation, where some X pronounced *Sem ga videl* as unique and tries to reconstruct what that person X could have had in mind by saying that, cryptophilology raises its head. The notion of context brings about a double risk. Cutting off the syntactic context short, one is left with lab sentences like Rus. *Мама мыла раму* 'Mom washed the frame', *Мама помыла раму, которую увидела* 'Mom washed the frame

---

6   Cf. Haspelmath (2014) for the discussion of referential scales and hierarchies.

she saw', etc. Expanding the context ad infinitum, one gets an open set of factors characteristic of unique speech situations and individual variation.

5. Texts, text corpora and variation

Texts collections are objects of a different type than individual texts. Modern text corpora often include information concerning the genre, date, and authorship of the texts they consist of, which gives a chance to retrieve linguistic, philological, and non-linguistic and non-philological information from the corpora, e.g. identify and locate the authors or measure the similarities of texts. I am not going to discuss the tools of corpus linguistics here, but I take issue with the idea that the format of the input data blurs the lines between language vs text. The language models predict what is well-formed and what is not. These predictions can, with few provisos, be tested on text corpora. On the one hand, all grammars are restrictive. On the other hand, variation is a normal state of all languages. The two sides of the problem are, whether a) the chosen model can be parametrized and tested, b) the chosen corpus is representative. If a corpus displays a regular contrast between several groups of speakers, e.g. part of the speakers accepts VS sentences in main clause declaratives, while the others do not, one deals with language-internal variation. The parametric description means that two or more idioms differ in the values of the same parameter.[7] Even codified languages like English or Russian are subject to "dialectal" divergences regarding some grammatical parameters (Wilson amd Henry 1998; Циммерлинг 2018). Restrictive rules, e.g. template rules predicting the contact position of Slavic clitics (Franks and King 2000; Зализняк 2008; Zimmerling and Kosta 2013) or word-building rules constraining the derivation of complex adjectives in Bambara (Vydrin 2018) tell licit combinations of clitics and adjectives from illicit ones. Both the positive prediction (what is expected to be found) and the negative one are relevant. The problem is that small and medium-sized corpora like Bambara corpus (Vydrin et al. 2011) do not always show all well-formed combinations, which could be elicited from native speakers and do not rule out all ill-formed combinations. Even large corpora, such as Russian National Corpus (RNC), prove small for certain tasks. The Russian word *тысяча* 'thousand' is morphologically a feminine noun, but in the subject position it displays some features typical for Russian cardinal numerals like *пять* '5', *девять* '9', *двадцать пять* '25', etc.: with the expressions from this class, the predicate either agrees in the plural – **пять человек** *пришли*$_{PL}$ 'five people came' – or assumes the default form of the singular neutrum, cf. **пять** человек *пришло* $_{SG.N}$. If one tries to learn from RNC the hybrid agreement of the word *тысяча*, one gets ca. 40 examples with the numeral phrase (NumP) **тысяча X-ов** 'thousand X-s' in the subject position. They show all three options – a) the predicate takes the plural form, cf. **тысяча** человек **пришли**$_{PL}$ '1000 people came'; b) the predicate stands in the singular neutrum, cf. **тысяча** человек **пришло**$_{SG.N}$; c) the predicate stands in the singular feminine, cf. **тысяча** человек **пришла**$_{SG.F}$. This sample does not tell us whether the same speaker really can select all three options in the same context. The phrases with an agreeing prepositive modifier (= Determiner phrase, DP) – cf. **эти**$_{PL}$ *тысяча X-ов* '**This** 1000 of X-s' and **эта** *тысяча X-ов* 'the same' – have different syntax: they only license the options a), c), and block

---

The same parametric approach can account for the cross-linguistic variation (Baker 2008; Лютикова and Циммерлинг 2018), but checking such models on parallel corpora is a far more demanding task technically.

the option b) *эти$_{\text{PL}}$ ***тысяча*** *человек* ***пришло***$_{\text{SG.N}}$, while phrases without an agreeing modifier license all three options.

The hypothesis that Russian grammar distinguishes the phrases of NumP-level and DP-level (Лютикова 2018) is theoretically non-neutral but is confirmed both by the elicited examples and the corpora. Language facts normally can be described in more than one way, but a valid alternative of a linguistic model predicting some distribution of units or structures is only a different model.[8]

6.   Cryptolinguistics in action

We concluded that irrespective of the fact whether cryptophilologists use innovative or mainstream terms, such analytic tools as text segmentation into fragments that correspond to the author's attitudes and his / her creative ideas place the research outside L. Hence follows that such end tasks as analysis of concepts specific to certain myths, cultures, or poetic traditions, likewise – author attribution, identification of speakers, and their linguistic portrayal belong to cryptophilology. At the same time, such end tasks as analysis of phonemic systems, preparation of grammatical descriptions, vocabularies and text corpora, identification of dialects, description of language diversity, and reconstruction of proto-languages belong to the core of L. There remains but one nasty question: what if a researcher knowingly or unknowingly combines linguistic and cryptophilological issues? Such cases are not infrequent. The same great mind, who argued in the 1960-s that there are no language units above sentences, Émile Benveniste around the same years wrote an essay on the etymology of the Indo-European word *peku 'sheep': he vigorously denied that I.E. *peku is connected with the root **pek-* 'cut wool' (or 'pluck wool') and asserted that the original meaning was 'movable personal possession' (Benveniste 1970). This is a purely linguistic hypothesis that can be right or wrong, cf. the reservations in Lehmann (1986: 63). But parallel to this, Benveniste also claimed that the Proto-Indo-Europeans had a cultural concept 'cattle-movables-wealth' which could be associated with *pek-u or other roots, sheep or other animals. It is unclear, whether this is a purely linguistic issue. There are no texts and text corpora for proto-languages. The usage of P.I.E. speakers regarding the concept 'movables' must be reconstructed through the examination of an open set of texts in different genetically related languages. Benveniste makes a brief survey of the situation in some Old I.E. languages and cites some formulaic expressions like Avestan *pasu vīra* 'cattle men' presumably characteristic of Old I.E. rites and habits. However, his second hypothesis requires a much deeper analysis of texts than the assessment of the connection between the stems **pek-* and *peku. One has to show the meaning structure of polysemous words and prove that the meanings like 'cattle', 'wealth' and 'money' are not only expressed by the same words but also are found in the same group of texts. This is a huge task even for a single well-documented language. Remarkably, Benveniste does not even follow the distribution of the stems **pek-* and *peku, being convinced that his intuition concerning the existence of the P.I.E. concept 'movables' is correct.

---

8   If a linguist dislikes constituent grammars and rejects the notion of generalized phrase structure applied by Lyutikova, (s)he still has to propose an alternative model, which explains the current distribution of the expressions like *эти тысяча X-в* vs *тысяча X-в* in Russian.

7. A Russian prince and a Wallachian boyar as L2 speakers

Cryptophilology is in the right place, where a given version of the text projects a set of related versions, cf. if a text from XIIth century survived in an XVIIth century copy, or if the author is an L2 speaker. For the speakers of Old Russian, Old Church Slavonic was a foreign, though related bookish language. It had to be learned. The great Russian scholar Alexei Sobolevsky asserted in 1890 with a directness shocking for our time that many authors from the XIth–XIVth centuries, including all chroniclers, Prince Vladimir Monomakh and the creator of the *Song of Igor's Campaign* had insufficient command of Old Church Slavic and produced hybrid texts (Соболевский 1980: 37):[9]

> They tried to write in Old Church Slavonic but were partly unable to express in Church Slavonic all that they wanted, partly did not know how to distinguish Church Slavonic from Russian. Therefore, we find a complete mixture of the Church Slavonic and Russian elements in their works.

Modern commentators are generally kinder: they proceed from the assumption that the hybrid texts from this group are written by people who had full command of Russian and discuss some linguistic parameters of Old Russian grammar, e.g. the distribution of enclitics as displayed in these texts (Зализняк 2008: 212-217).[10] A linguistic approach like this is a valid alternative to Sobolevskij's declaration that the mixture of Church Slavonic and Russian in the aforementioned group of vernacular texts was *complete.* Still, cryptophilological issues cannot be ignored here, even if one is primarily interested in language features. There are different opinions concerning the written heritage of Monomakh, let us adopt the view that all texts ascribed to this author in the Laurentius chronicle and placed there under 1096 A.D. were supervised and compiled by him (Гиппиус 2003). Monomakh mostly uses Church Slavonic simple past forms (aorist and imperfect), while colloquial Old Russian as we now know after the discovery of the Birchbark letters, lacked them and replaced them with the perfect (narrative) construction with l-participle. However, towards the end of two sections – the conclusion of Monomakh's autobiography and his dramatic letter to his enemy Prince Oleg – the ratio of the perfect drastically increases. This might be explained by the fact that the corresponding contexts favor the use of the perfect, but similar contexts are found in the preceding parts too, and in some cases, where Monomakh used the perfect, one could also expect aorist or imperfect, cf. Sobolevskij's view that Monomakh lacked sufficient control of Church Slavonic. It is therefore not excluded that in some poignant situations as in the final section of his life summary and the conclusion of the message to Oleg, where he tells that Oleg still holds two of his young sons as hostages after having killed his grown-up son, Monomakh switched the code from Church Slavonic to Russian. The cryptophilological perspective looks promising: it hints that depicting emotionally-loaded situations, Russian authors from the XIIth century could revert to their L1 grammar, which was normally suppressed in the narrative fragments of

---

[9] This book by Sobolevskij was not published in his lifetime.

[10] An anonymous reviewer wonders whether this formulation is applicable to another book of Zaliznjak (2007), where he proves that the text *Song of Igor's Campaign* was not forged, since it displays the Old Russian grammar of the late XIIth century. There are two relevant aspects. First, there is a substantial body of Old Russian texts from the XIIth century, therefore one can check whether the grammar of spurious texts is compatible with the descriptions of available original texts. Second, author distribution, reconstruction of the protograph and identification of later insertions are cryptophilological tasks.

their books. Meanwhile, the linguistic implications of the cryptophilogical analysis of Monomakh's text are thin: we just confirmed our guess that he could use the perfect construction as an L1 speaker. To get less trivial results, a linguist must take more language parameters, e.g., consider the distribution of clitics and word orders in the Old Russian perfect construction.

A slightly different situation is attested in the corpus of Wallachian letters (1386–1506 A.D.). These letters are written by L2 speakers of Middle Bulgarian, who also spoke Rumanian. Middle Bulgarian was inter alia used as *lingua franca* in the correspondence of Wallachians with the multilingual community of the German city of Braşov (alias Kronstadt). Middle Bulgarian pronominal and auxiliary clitics are normally placed clause-internally and behave as enclitics: they are not fronted and do not attach to initial proclitics. A remarkable exception comes from one letter of the Wallachian boyar Gherghe Lascar (1437), who put the clustering clitics after the proclitic 6 times, cf. (1a–b). This can be a feature of his first language, Rumanian. There is, however, an alternative explanation. The placement of the enclitics after the proclitics can be a means of emphasis as in Modern Bulgarian, where part of the speakers accepts such sentences in emphatic contexts (Iliyana Krapova, p.c.). Gherghe indeed expresses his anger by the fact that his trading partner *Hanesh*, a citizen of Braşov, deceived him, and the magistrates of Braşhov did not help. In the end, he says 'my heart is angry', cf. (2).

(1)   a.   **И =ми** поръчи съ Ханеш
              'And he sent me a message with Hanesh.' (Gherghe Lascar, l. CCXIII 1437)
       b.   **И =смь** писал вам, да не кажете утрѣ или в други день : нѣ ни си писал, да не знамо ми ѡ тѡм.
              'And after all, I wrote to you so that you would not say tomorrow or another day "No, you did not write that, we do not know that".' (Gherghe Lascar, l. CCXIII, ca. 1437)

(2)   Тогизи да се не злосръдите, понеже зло ест мое сръдце.
       'Do not be angry at this because my heart is <very> angry.'
       (Gherghe Lascar, l. CCXIII, ca. 1437).

Although Gherghe does not use gross words here, the tone of his letter is aggressive and the proclitic-enclitic sequences *И=ми порѫчи, И=смь казал*, etc. probably mark his anger. Unfortunately, the data is short to prove that such word order was a means of emphasis.[11] Neither can one prove the alternative guess that the placement of enclitics after initial proclitics was an L1 feature of the Wallachian speakers of Middle Bulgarian suppressed by the L2 grammar. The best solution for a linguist primarily interested in describing Middle Bulgarian word order is to cut off all factors of Gherghe's clash with Hanesh in 1437 and state that Middle Bulgarian clustering pronouns and auxiliaries are normally placed after clause-initial stressed elements, apart from anomalous contexts, where they attach to proclitics. But if a researcher is primarily interested in reconstructing the emotions of medieval people (cryptophilology) or the bilingual situation in Wallachia (probably L), or compares languages with clitic fronting (mainstream L), the answer to the question, what underlined the choice of the word order (1a–b) is relevant.

---

[11]   It is but noteworthy that in Gherghe Lascar's second letter, which is more quiet, since his trading partner finally arrived – *А нинѣ ест дошел Ханеш до мене* 'And now Hanesh came to me' (l. CCXIV, ca. 1437) there are no proclitic-enclitic sequences.

8. Conclusions

We followed the contrast of linguistics and cryptophilology. The latter borrows from traditional philology the non-linguistic concept of text as fixed verbal form, but claims the right to analyze it as a linguistic object. The demarcation principle is straightforward. There are no language units larger than complex sentences, the text segmentation from above into fragments lacking syntactic structure is not a linguistic issue. Linguists apply to the notion of context to get a distribution of language units or structures but must cut off the context as soon as this task is solved. Cryptophilologists aim at explaining the unique features of a text and the situation where it has been verbalized, in order to get a correlation between text fragments and the author's intentions or feelings.

References

Aissen Judith (2003): "Differential object marking: Iconicity vs economy," *Natural Language and Linguistic Theory*, 21, pp. 435–483.

Austin John L. (1962): *How to Do Things with Words: The William James Lectures delivered at Harvard University*, Oxford, Clarendon Press.

Baker Mark (2008): "The macroparameter in a microparametric world," in T. Biberauer ed., *The limits of variation*, Amsterdam / John Benjamins, pp. 351–374.

Bally Charles (1932): *Linguistique générale et linguistique française*, Paris, Leroux.

Benveniste Émile (1962): "Les niveaux de l'analyse linguistique," in *Preprints of Papers for the Ninth International Congress of Linguists*, Cambridge, Mass.

Benveniste Émile (1970): "Les valeurs économique dans le vocabulaire indo-européen // Indo-European and Indo-Europeans," in *Papers presented at the 3ʳᵈ Indo-European conference at the Univ. of Pennsylvania*, Philhadelhia, pp. 307–310.

Daneš František (1974): "Functional sentence perspective and the organization of the text," in F. Daneš ed., *Papers on functional sentence perspective*, Prague, Academia, pp. 106–128.

Franks Steven L. and King Tracy H. (2000): *A Handbook of Slavic Clitics*, Oxford, Oxford University Press.

Grice Herbert Paul (1975): "Logic and Conversation," in P. Cole and J. Morgan, *Syntax and Semantics*, vol. 3, New York, Academic Press, pp. 41–58.

Haspelmath Martin (2014): "Descriptive Scales vs Comparative Scales," in I. Bornkessel-Schlesewsky, A. L. Malchukov, and M. D. Richards eds., *Scales and hierarchies*, Berlin, De Gruyter Mouton, pp. 45–58.

Heusler Andreas (1923): *Die altgermanische Dichtung*, Berlin, Athenaion.

Heusler Andreas (1969): *Kleine Schriften* 1–2, Berlin, De Gruyter.

Jakobson Roman (1973): "Postscriptum," in *Questions de Poétique*, Paris, Sculle, pp. 485–504.

Lambrecht Knud (1994): *Information structure and sentence form. Topic, focus and the mental representation of discourse referents*, Cambridge, Cambridge University Press.

Lehmann Winfred P. (1986): *A Gothic etymological dictionary*, Leiden, Brill.

Mᴀᴛʜᴇsɪᴜs Vilém (1939): "O tak zvanem aktualnim cleneni vetnem," *Slovo a Slovnost*, 5, pp. 171–4.

Mᴇʟ'čᴜᴋ Igor (2016): *Language: From Meaning to Text*, Ars Rossica series, Moscow & Boston, Academic Studies Press.

Sᴀssᴇ Hans-Jürgen (1987): "The thetic / categorical distinction revisited," *Linguistics*, 25:3, pp. 511–580.

Sɪᴅᴏʀᴏᴠ Grigory, Kᴏʙᴏᴢᴇᴠᴀ Irina M., Zɪᴍᴍᴇʀʟɪɴɢ Anton, Cʜᴀɴᴏɴᴀ-Hᴇʀɴáɴᴅᴇᴢ Liliana, and Kᴏʟᴇsɴɪᴋᴏᴠᴀ Olga (2014): "Modelo Computacional del Diálogo Basado en Reglas Aplicado A Un Robot Guía Móvil," *Polibits*, 50, pp. 35–42.

Sᴇᴀʀʟᴇ John (1982): *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press.

Vʏᴅʀɪɴ Valentin (2018): "Where corpus methods hit their limits: The case of separable adjectives in Bambara," *Рема. Rhema*, 4, pp. 34–49.

Vʏᴅʀɪɴ Valentin, Mᴀsʟɪɴsᴋʏ Kirill, Méʀɪᴄ Jean-Jacques, and Rᴏᴠᴇɴᴄʜᴀᴋ Andrij (2011): *Corpus Bambara de Référence*, ᴜʀʟ: http://cormand.huma-num.fr/index.html (visited 07/06/2022).

Wɪʟsᴏɴ John and Hᴇɴʀʏ Alison (1998): "Parameter setting within a socially realistic linguistics," *Language in Society*, 27:1, pp. 1–21.

Zɪᴍᴍᴇʀʟɪɴɢ Anton and Kᴏsᴛᴀ Peter (2013): "Slavic clitics. A typology," *Sprachtypologie und Universalforschung (STUF)*, 66:2, pp. 178–214.

Бʟᴏх Марк Я. (2000): «Диктема в уровневой организации языка», *Вопросы языкознания*, 4, с. 56–67.

Гɪᴘᴘɪᴜs Алексей А. (1996): «"Русская Правда" и "Вопрошание Кириково" по Новгородской Кормчей 1282 г.: к характеристике языковой ситуации Древнего Новгорода», *Славяноведение*, с. 48–62.

Гɪᴘᴘɪᴜs Алексей А. (2003): «Сочинения Владимира Мономаха. Опыт текстологической реконструкции. I», *Русский Язык в научном освещении*, 2, с. 60–99.

Зᴀʟɪᴢɴʏᴀᴋ Андрей А. (2007): *«Слово о Полку Игореве». Взгляд лингвиста*, М., ЯСК.

Зᴀʟɪᴢɴʏᴀᴋ Андрей А. (2008): *Древнерусские энклитики*, М., ЯСК.

Кᴏᴠᴛᴜɴᴏᴠᴀ Ирина И. (1976): *Современный русский язык. Порядок слов и актуальное Членение Предложения*, М., Просвещение.

Лʏᴛɪᴋᴏᴠᴀ Екатерина А. (2018): *Структура именной группы в безартиклевом языке*, М., ЯСК.

Лʏᴛɪᴋᴏᴠᴀ Екатерина А., Цɪᴍᴍᴇʀʟɪɴɢ Антон В. (2018): «Почему языки такие предсказуемые? Типология морфосинтаксических параметров», *Типология Морфосинтаксических параметров*, Т. 1 (2018), Вып. 1, с. 11–30.

Пᴀᴅᴜᴄʜᴇᴠᴀ Елена В. (1996): *Семантические исследования. Семантика времени и вида в русском языке. Семантика Нарратива*, М., ЯСК.

Пᴀᴅᴜᴄʜᴇᴠᴀ Елена В. (2018): *Эгоцентрические единицы языка*, М., ЯСК.

Сɪʙɪʀʏᴀᴋᴏᴠ Андрей В. (2020): *Текстемная организация специальных отраслевых Текстов (На материале русских и английских текстов по виноделию)*. Автореферат диссертации ... кандидата филологических наук, М.

Сᴍɪʀɴɪᴄᴋᴀʏᴀ Ольга А. (1994): *Язык древнегерманской поэзии*, М., МГУ.

Соболевский Алексей И. (1980): *История русского литературного языка*, Л., Нау-
    ка.

Стеблин-Каменский Михаил И. (2003): *Труды по филологии*, СПБ., СПБГУ.

Циммерлинг Антон В. (2018): «Два Диалекта русской грамматики: корпусные дан-
    ные и модель», *Компьютерная лингвистика и интеллектуальные технологии*,
    Вып. 17 (24), с. 818–830.

Циммерлинг Антон В. (2021): *От интегрального к аспективному*, М., ЯСК.

Янко Татьяна Е. (2008): *Интонационные стратегии русской речи*, М., ЯСК.

Национальный Корпус Русского Языка, URL: https://Ruscorpora.ru/.