ПЛОВДИВСКИ УНИВЕРСИТЕТ „ПАИСИЙ ХИЛЕНДАРСКИ“ – БЪЛГАРИЯ
НАУЧНИ ТРУДОВЕ, ТОМ 61, КН. 1, СБ. Б, 2023 – ФИЛОЛОГИЯ,
PAISII HILENDARSKI UNIVERSITY OF PLOVDIV – BULGARIA
RESEARCH PAPERS, VOL. 61, BOOK 1, PART B, 2023 – LANGUAGES AND LITERATURE

DOI 10.69085/ntf2024b242

DEVELOPMENT TRENDS IN THE MORPHOLOGICAL-TYPOLOGICAL PROFILE OF BULGARIAN. A QUANTITATIVE PERSPECTIVE

Nikolay Stankov
St. Cyril and St. Methodius University of Veliko Tarnovo

The study investigates from a quantitative perspective the overall development of the morphology of the Bulgarian language in the past 100 years. It aims to determine whether and to what extent the language has become more or less analytic or synthetic. Additionally, the quantitative morphological differences between three text types – literary, scientific and journalistic – are measured and analysed. The study finds that Bulgarian has simultaneously become less analytic and less synthetic, resulting in a decrease of the grammatical information encoded in texts. A high level of syntheticity is measured in scientific texts, while literary works exhibit a relatively high level of analyticity.

Key words: typological profiling, analyticity index, syntheticity index, grammaticity, Bulgarian

1. Introduction

The rapid development of information and computing technologies and their growing capability to analyse large quantities of data swiftly and efficiently has enabled linguists in the past two decades to rely more and more on empirical and quantitative methodologies in their research. In addition, the novel technologies are also simultaneously opening up entire new fields of study with the emergence of artificial intelligence, machine translation and various other forms of natural language processing by computers. Nevertheless, older fields of linguistic research, like phonology, morphology, typology etc. can also profit substantially from the tools offered by powerful computers and the world wide web. Morphological typology lies in the foundation of what is nowadays considered linguistic typology. Pioneers in the field like A.W. Schlegel, W. von Humboldt and A. Schleicher were first intrigued precisely by the morphological differences between the ancient languages and the Indo-European languages of the 19th

century. The classification of languages according to particular criteria as well as the examination of their diachronic development has been going on for more than 150 years now, but it seems a large proportion of what has been done, especially in the early stages of the field's development, has been based on a qualitative research approach.

While qualitative research has yielded and continues to yield many valuable insights into the workings of linguistic systems around the world, quantitative research has the potential to supply empirical evidence and an increased level of objectivity to observations and conclusions that were originally based on rather subjective intuitions. Over the years, a number of linguists have noticed that, in their historical development, many Indo-European languages, including English, tend to abandon synthetic morphological structures, that encode grammatical information with the help of endings or vowel mutation, among other things, in favour of more analytic structures, that transmit grammatical information by using free grammatical markers, like auxiliary verbs, pronouns etc. (cf. Schlegel 1818: 17; Davenport 1992: 27; Siemund 2004: 169; Tristram 2009: 255; Haspelmath and Michaelis 2017: 2; Uktamovna 2023: 46). While the general analytisation of Indo-European languages seems to have established itself as a widely accepted fact, quantitative evidence could be very helpful in determining whether this is indeed correct and if so, what are the characteristics of this process. The present study aims to collect and evaluate precisely this type of evidence.

Nevertheless, it should be noted here that the idea of a process of general analytisation in the Indo-European languages has been challenged on multiple occasions in the recent years, as evidence for the emergence of new synthetic constructions in some languages in the family started to surface (cf. Ledgeway 2011: 384f.). In addition, the classification of languages within the framework of the analytic-synthetic continuum has also been criticised (cf. Schwegler 1990: xii; Hollmann 2009: 548f.), with some suggesting only linguistic constructions should be described in this way (cf. Hollmann 2009: 550). In addition, to the extent that it can be spoken of analytic or synthetic shifts within a language, it has been proposed that the morphological system of the Indo-European languages is not moving strictly in only one linear direction (i.e. towards greater analyticity or towards greater syntheticity), but instead its change is cyclic, i.e. there have been many interchanging periods of high analyticity and high syntheticity (cf. Hollmann 2009: 550; Szmrecsányi 2016: 93).

2. Aims and methodology of the study

2.1. Aims

The present study's goal is to create a quantitative morphological-typological profile of modern Bulgarian by measuring the levels of analyticity, syntheticity and the related metric of grammaticity of the morphological system of the language in a quantitative way across three distinct text types (literary, scientific and journalistic) by using corpus data. The differences between the individual text types are also analysed and discussed. Another aim is to track how the levels of analyticity, syntheticity and grammaticity of the language have changed diachronically over the past 100 years. In addition, after analysing the data, I attempt to provide possible reasons and explanations for the results. The same methodology was recently applied in a very similar study of German (cf. Stankov 2024, forthcoming). While a direct comparison between the two languages is envisioned for future works, some parallels between the two datasets are drawn here too. Since the Indo-European language family is one of the most widely spread around the globe, the same methodology could be adapted and applied to other languages within the family and eventually to languages from other language families in order to compare their morphological profiles and track the direction of their development.

2.2. Corpus composition and method of analysis

The methodological basis for the current study comes from Szmrecsányi and Kortmann (2011), Szmrecsányi (2016) and Horsch (2021), which are in turn originally based on Greenberg (1960). However, it has to be pointed out that some changes have been introduced to the aforementioned researchers' approach in order to achieve a more complete account of the morphological system of the analysed language. As already mentioned, the adapted method has already been used once for a similar analysis of German texts (cf. Stankov, forthcoming). The main source of material for the analysis here are six corpora of approx. 50,000 word tokens each. The corpora were compiled manually and specifically for the purpose of the study from texts available on the Internet. This approach was preferred to using ready-made online corpora because free access to the text of the corpus was essential for the study, as the analysis is based on randomly selected samples from the corpora. In addition, by compiling the corpora manually, I was able to hand-pick the texts included in the corpora and ensure they match several important criteria: their time of publication, the text type to which they belong and the language in which they were written

originally (especially in the corpus of literary works attention was paid to make sure the texts were written by Bulgarian authors).

Three out of the six corpora contain texts from the period 1918-1922, while the other three comprise texts from the period 2018-2022. Within each of these two groups of three, one corpus contains literary texts (excerpts from novels by Bulgarian authors that were originally published within the respective period), one includes scientific texts (excerpts from scholarly and academic works published within the respective period) and one encompasses journalistic texts (articles from digitised newspapers and news websites from the respective period). Similarly to what was done previously for German (cf. Stankov, forthcoming), five thousand token samples were extracted from each of the corpora described here. The selection process was randomised with the help of a random number generator, which identified the section of corpus text to be extracted. Herein lies one of the differences from the approach taken by Szmrecsányi and Kortmann (2011: 173), who also used a randomised selection process, but in their version every single word in the sample is randomly selected. Apparently, this approach could only yield samples with incoherent text, which I find less representative of the actual state of the language than a sample of naturally composed, coherent text.

After selection and extraction each sample was run through a POS tagger software, called TagAnt, created by Lawrence Anthony (2022). This tool is based on the engine of the text annotating software TreeTagger by Schmid (1994; 1995). The tagset used for Bulgarian is described in detail by Simov, Osenova and Slavcheva (2004). TagAnt assisted the analysis by identifying the part-of-speech class of every word token in the analysed samples. Once processed, the tagged samples were examined manually, and the elements that contained grammatical meaning were categorised either as analytic tokens or as synthetic tokens. The following parts of speech were allocated to the analytic token class: pronouns (including personal, demonstrative, collective, possessive, interrogative, relative, negative and indefinite pronouns), auxiliary verbs, modal verbs (including *мога* ‘can’, *трябва* ‘must/should’ and *искам* ‘want’), conjunctions, subjunctions, prepositions, some particles (including negative, interrogative, auxiliary and gradable) and existential uses of the verbs *има* ‘there is’ and *няма* ‘there isn’t’. After the manual examination and the categorisation of the tokens, the number of analytic tokens in each sample was counted. The number of analytic tokens per 1,000 words is called the analyticity index (AI). A mean analyticity index (mean AI) was determined for each corpus by combining

the AI of every sample extracted from that corpus and dividing the sum by the number of samples.

In addition to the free grammatical markers listed above, which constitute the analyticity index, bound grammatical markers that encode grammatical meaning within the boundaries of the orthographically undivided word token were also identified and counted. The following types of bound markers were counted as synthetic tokens: case, number and grammatical gender markings on nouns, verb participles, pronouns, adjectives, adverbs and numerals (e.g. *-и* in *врати* 'doors' or *-а* in *негова* 'his'); verb conjugation inflexions expressing tense, person and number (e.g. *-ше* in *бягаше* 'he/she/it ran' or *-ме* in *скачаме* 'we jump'), including consonant mutation or the addition of elements to signal aspectual differences (e.g. the alternation between *-зв* and *-ж* in the verb couple *казвам/кажа* 'I am saying/to say' or the addition of *-в* in *чувам* 'I hear/I am hearing', distinguishing it from *чуя* 'to hear'); grammatical markers deriving the various verb participles from their basic forms (e.g. *-н* in *видян* 'seen', derived from *видя* 'to see' or *-ейки* in *ходейки* 'walking', derived from *ходя* 'walk'); the definite article (e.g. *-ът* in *човекът* 'the man' or *-та* in *масата* 'the table'). The number of all bound grammatical markers in the thousand-token sample is its syntheticity index (SI)[1]. A mean syntheticity index (mean SI) was determined for each corpus by combining the SI of every sample extracted from that corpus and dividing the sum by the number of samples.

In addition to the two already described indices, a third index was calculated by combining the mean AI and the mean SI of every corpus. This grammaticity index (GI) represents the total number of overt grammatical markers (bound or free) per 1,000 words. However, a GI was not calculated for every sample, as it is above all, the mean grammaticity index (mean GI) that is of value for the discussion, and this could simply be calculated by adding up the mean AI and the mean GI of a corpus. The term *grammaticity* will be briefly discussed below, as it is not as common as analyticity and syntheticity. The indices of the samples gained from the corpora comprising 21st century texts as well as the mean indices (mean AI, mean SI and mean GI) can be found in Table 1. The same kind of data about the samples gained from the corpora comprising 20th century texts is available in Table 2.

[1] Note that in case of multiple bound grammatical markers expressing the same information, they are counted as one single syntheticity token.

2.3. Corpus size remarks

In the study covering German (cf. Stankov, forthcoming), I argue that a corpus size of 50,000 tokens may appear quite insignificant when compared to modern corpora available online for free, but it should be sufficient for the purposes of the present kind of study. In summary, the main arguments in favour of this statement are the following: firstly, each corpus is designed with a particular text type and historical period in mind; secondly, the corpora are not the subject of the analysis, but only the samples that are extracted from them; and lastly, online corpora do not always offer free access to their entire text, which makes the extraction of samples impossible. They are, therefore, incompatible with the design of the study. Since the present study has the same design as the one examining German, the reader is advised to consult the aforementioned publication for a more detailed discussion of corpus size.

2.4. Concept definitions

This paper makes use of a number of concepts that should be defined in the context of the study, as they could be construed in different ways. These include: *analyticity*, *syntheticity*, *grammaticity*, *text type*, *literary*, *scientific* and *journalistic text* and *morphological-typological profile*. Since the definitions of these terms have already been discussed in the context of the study dealing with German, here I will only briefly expound on the two more rarely used concepts: *morphological-typological profile* and *grammaticity*. For a discussion of the other terms, the reader may review Stankov (forthcoming).

The concept of a *morphological-typological profile* can generally be understood as describing the morphological characteristics of a natural language system in the broad context of linguistic typology. While this is a useful definition, it must be pointed out that the present study aims to produce a quantitative morphological-typological profile and not a qualitative one. The latter would be a much larger undertaking than the scope of this paper allows. Therefore, in the current context the term should be understood as denoting the level of analyticity and syntheticity of the morphological system of a language. As can be seen, the concepts of analyticity and syntheticity play a central role in the definition of a quantitative morphological-typological profile. I believe this is acceptable for the following reasons: firstly, the two concepts are simultaneously sufficiently broad (i.e. they cover a number of easily identifiable linguistic features that a language may or may not possess, e.g. case markings or auxiliary verbs) and sufficiently distinguishing (i.e. the linguistic features

they incorporate are significantly different from one another, e.g. using prepositions to express relations between clause constituents is a significantly different strategy from using case markings). Secondly, the two concepts are widely accepted and well-known in the academic community, although it must be admitted that they have also been the object of a lot of criticism, especially in more recent years. Thirdly, although abstract in nature, the two concepts can be quantified, as, for example, Greenberg (1960) shows.

The other term that we need to explain here, as it may not be immediately familiar to the reader, is *grammaticity*. Here, this concept is borrowed from Szmrecsányi (2009: 322). He describes it as comprising "all explicit grammatical markers, but not word order" and "a ratio of the total number of grammatical markers [...] in a text to the total number of words". Essentially, the concept of grammaticity measures the amount of grammatical information encoded in a text. For this reason it can be a valuable instrument in the study of the morphological-typological profile of a language.

3.Corpus data analysis

3.1. Literary text type

We begin the analysis of the data with the literary text type. The mean AI for the 21st century literary corpus amounts to 411 and the mean SI to 525 (see Table 1). Comparing these results to those of the other two corpora for the same historical period, we see that this corpus exhibits the highest mean AI score within this corpus group. Interestingly, a very similar observation was made in the parallel study of German (cf. Stankov, forthcoming). In it, I attributed the high analyticity, measured in the samples from this corpus, to "the presence of dialogue in narrative-driven literary texts" (Stankov, forthcoming). Of course, an additional investigation of oral speech would have to be conducted, before this hypothesis could be confirmed. Alternatively, a more in-depth part-of-speech analysis of the literary corpora used here could also help discern whether the high level of analyticity is attributable to the imitation of everyday language, found in many fictional literary works. In any case, the measurements made in the present study support those from the study of German, because, though the two languages may be typologically different, the literary text type in both languages exhibits similar characteristics.

Taking a look at the results from the 20th century corpora, we discover a similar pattern: with a mean AI of 444 and a mean SI of 586, the literary

corpus has the highest mean AI in this group of corpora. However, it is also noticeable that both scores are higher than their 21st century counterparts. This means that over the course of the past 100 years literary texts have simultaneously become less analytic and less synthetic. As we will see below, this is a pattern that is also observed in the other corpora and indeed seems characteristic of the entire language system. Since both mean AI and mean SI affect the mean GI score, not surprisingly, the mean GI for the 21st century literary corpus (936) is lower than its 20th century counterpart (1030). In line with the explanation of the concept of grammaticity, provided above, this can be interpreted as a decline in the amount of grammatical information that Bulgarian literary texts encode.

3.2. Scientific text type

The examination of the two corpora containing scientific texts also leads to some interesting, and perhaps anticipated, observations. In both the 21st and 20th century groups of corpora, the corpora containing scientific texts have the highest mean SI, compared to the other text types. Again, a similar observation was made in the study covering German (cf. Stankov, forthcoming), where scientific texts also exhibited the highest level of syntheticity. In that study, I suggest the high mean SI of scientific texts may be a consequence of their focus on detailed descriptions and explanations, which often involve the increased use of adjectives and nouns. In Bulgarian, these are some of the parts of speech that most often receive inflections, which is a synthetic strategy for the encoding of grammatical information. In addition, I have argued that there is a general tendency in academic texts to use a more sophisticated style of writing, which features heavier noun and verb phrases and consequently more cases of adding inflections. At the same time, scientific texts are generally unlikely to feature dialogues and everyday language. Just like in German, however, here too, a more comprehensive part-of-speech analysis will be necessary to test if and to what extent this explanation is viable.

While the 21st century scientific corpus exhibits the highest level of syntheticity, compared to the other two corpora from the same period, it also has the lowest mean analyticity index. Indeed, taking a look at the mean GI, we notice that the scores for the literary and scientific corpora are even at 936. Taken alone, these two observations suggest that there may be a trade-off between syntheticity and analyticity and that the total amount of grammatical information encoded in the different text types is virtually the same. However, the mean GI of the 21st century journalistic corpus tells a different story, as it reaches only 893. This difference between the mean GI

of literary and scientific corpora, on the one hand, and the mean GI of the journalistic corpus, on the other, is even more pronounced in the 20th century group. This discrepancy indicates that, even across different text types within the same language system, there are differences in the amount of grammatical information that texts carry, although on average these differences may be small.

Speaking of the results from the 20th century corpora, it is worth noticing that, similarly to what was observed in the literary corpora, here too we can recognise a decline in both analyticity and syntheticity for the scientific texts (see Table 1 and Table 2). However, the decrease in analyticity in the scientific texts (65 index points) is stronger than the decrease in analyticity in the literary texts (33 index points), whereas, when it comes to syntheticity, the opposite development is the case: the decrease in syntheticity in scientific texts (39 index points) is smaller than the decrease in syntheticity in literary texts (61 index points). While it is difficult to assess what are the precise factors influencing the speed of the decline in analyticity and syntheticity in these two text types, I would venture to speculate that scientific texts are slower to become less synthetic precisely because of the more linguistically conservative and rigid nature of scientific discourse and the preference for more complex expressions and structures, which, at least in Bulgarian, often involve higher usage of inflections. On the other hand, literary texts may lose syntheticity faster because of the relative freedom that authors of fiction enjoy when it comes to their choice of style and expression. In any case, whatever the reasons may be, from the current dataset it is evident enough that both metrics are registering a decline, which in itself is an insightful observation, as it suggests that analyticity and syntheticity are not necessarily two concepts constituting a continuum, and a decline in one metric does not necessarily entail an increase in the other, as a continuum situation would implicate.

3.3. Journalistic text type

The 21st century journalistic corpus has a medium to low mean AI (377) and a comparatively low mean SI (516). The low analyticity in journalistic texts is something that was also observed in the parallel study dealing with German (cf. Stankov, forthcoming). There, I speculated that this may be due to the absence of everyday language and dialogue in this

type of text[2] and to a general preference for brief and economic expressions, designed to save reading time and printing space. The low syntheticity in these texts, on the other hand, may be attributable to the fact that these texts are not as focused on detailed descriptions and explanations as the other two text types. The aspiration to save time and space in modern day news reports and analyses may also be reflected in the mean GI score of 893, which is noticeably lower than that of the other two corpora. It appears as though this text type has been shaking off some of what can be considered unnecessary or expendable grammatical information or "ornament" as McWhorter (2001: 410 as cited in Kortmann and Szmrecsányi 2009: 267) calls it.

The results from Table 2 tell a similar yet slightly different story. In the group of 20th century corpora, the journalistic corpus exhibits both the lowest mean AI (393) and the lowest mean SI (536). However, both scores are higher than their 21st century counterparts. The difference is not big, though: 16 index points for mean AI and 20 for mean SI. A directionally similar but stronger decline in both metrics was registered in the journalistic corpora for German (cf. Stankov, forthcoming). While the decrease in Bulgarian is smaller, it is still an indicator for ongoing developments in the journalistic writing style that have occurred during the 20th and 21st centuries. In the case of German, I attribute this to what I perceive as a higher tendency for political commentary and ideological bias in the texts from the 20th century journalistic corpus and a greater focus on neutrality and matter-of-fact informativeness in the 21st century journalistic texts. For the Bulgarian counterparts I made very similar observations while processing the samples. The proposed explanation is of course a very tentative hypothesis that will have to be substantiated by a closer analysis of the texts from the respective corpora.

3.4. Comparison of overall results

In linguistic circles Bulgarian is often portrayed as an analytic language, especially in comparison to other Slavic languages (cf. Hinrichs 2004: 381, Osenova 2010: 644, Levshina 2020: 73, Banasiak 2021: 1). While the comparatively high analyticity of Bulgarian among its genetic relatives is hard to deny, the results from the present study indicate that the Bulgarian morphology may still retain a relatively high level of syntheticity (overall mean SI for 21st century Bulgarian texts is 538) for a language that

2 Note that the journalistic corpora in the present study do not include interviews, which are journalistic texts that do include dialogue. The inclusion of interviews in the analysis could therefore alter the end result for this text type.

is viewed as an analytic one[3]. The overall mean SI of the 21st century corpora exceeds their overall mean AI by 154 index points, and this is no small gap. In the parallel study covering German, the difference between overall mean SI and overall mean AI for the 21st century texts is only 14 index points (cf. Stankov, forthcoming). This shows that German relies on a more balanced distribution of encoding strategies between analytism and syntheticism, while Bulgarian uses synthetic strategies more heavily.

In the 20th century corpora for Bulgarian, the difference between the two overall metrics is similar – 156 index points. This observation suggests that the distribution of encoding strategies for grammatical information has not changed significantly in the past 100 years. However, if we compare the overall mean AI scores and the overall mean SI scores of the two groups of corpora, instead of the differences between them, it becomes easy to recognise that both metrics are lower in the 21st century corpora. The difference in overall mean AI is 38 index points, while the difference in overall mean SI is 40 index points. The results are similar, but not insignificant. Namely, they demonstrate that the morphology of the Bulgarian language has simultaneously become less analytic and less synthetic in the course of the last 100 years. As a result, the amount of grammatical information that texts in Bulgarian encode via morphological means has decreased. This is evidenced by a comparison of the overall mean GI scores of the two groups of corpora, which reveals a difference of 78 index points. In other words, this is a 7.8% decrease over a period of 100 years. This is an intriguing finding, but caution should be exercised before forming further reaching conclusions. An investigation of a different kind of data (as opposed to the corpus data used in the present study), such as survey data, and an examination of individual morphological features (cf. for example Kortmann and Szmrecsányi 2009: 270) in Bulgarian could be useful to test and prove the claims of the present paper. In addition, an analysis of Bulgarian syntactic structures may also prove fruitful, as the decrease in grammatical information encoded morphologically could be compensated for by a proliferation of syntactic encoding strategies.

4. Conclusion

The present quantitative study provides some general insight into the current level of morphological analyticity and syntheticity of the Bulgarian language and into the development of the language with regard to these two metrics over the past 100 years. However, qualitative research will be

3 For comparison, the overall mean SI for 21st century German texts is 446.

necessary to solidify and further elucidate the trends discovered here. Expanding the research field of the methodology employed in this study to other languages and historical periods could yield more valuable understanding into the general direction and speed of typological change in the major linguistic families of the world. In particular, as a next step, it would be interesting to see whether other Indo-European languages have also registered a decline in both morphological analyticity and syntheticity over the past century. Furthermore, looking at older periods and development stages of the languages could reveal whether this trend has been going on for a long time and whether morphological-typological change is indeed cyclical, as some have suggested (cf. Szmrecsányi 2016). This kind of research could also contribute towards measuring and analysing the development of the complexity of human language (cf. Kortmann and Szmrecsányi 2009). As a natural and almost inseparable part of human life, languages and their change over time could reveal much about the nature and history of humankind.

Table 1

Corpus	**Sample 1**		**Sample 2**		**Sample 3**		**Sample 4**		**Sample 5**				
	AI	**SI**	**AI**	**SI**	**AI**	**SI**	**AI**	**SI**	**AI**	**SI**	**Mean AI**	**Mean SI**	**Mean GI**
Literary corpus 21 Cent.	319	507	411	537	432	456	479	528	415	597	**411**	**525**	936
Scientific corpus 21 Cent.	363	521	361	557	366	642	312	533	345	586	**363**	**573**	936
Journalistic corpus 21 Cent.	369	536	301	378	376	569	426	563	411	537	**377**	**516**	893
Overall											**384**	**538**	**922**

Table 2

Corpus	**Sample 1**		**Sample 2**		**Sample 3**		**Sample 4**		**Sample 5**				
	AI	**SI**	**AI**	**SI**	**AI**	**SI**	**AI**	**SI**	**AI**	**SI**	**Mean AI**	**Mean SI**	**Mean GI**
Literary corpus 20 Cent.	406	635	445	614	470	499	432	598	468	585	**444**	**586**	1030
Scientific corpus 20 Cent.	381	700	441	604	427	562	448	592	442	601	**428**	**612**	1040
Journalistic corpus 20 Cent.	369	471	310	507	450	619	400	547	437	537	**393**	**536**	929
Overall											**422**	**578**	**1000**

REFERENCES

Anthony 2022: Anthony, L. *TagAnt* (Version 1.3.0) [Computer Software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software/tagant/> (accessed Sept. 2022)

Banasiak 2021: Banasiak, J. Built-in Argument Positions in Bulgarian and Polish. // *Cognitive Studies | Études cognitives*, Vol. 2021, No. 21, Article 2558. <https://doi.org/10.11649/cs.2558> (accessed Sept. 2023)

Davenport 1992: Davenport, P. The De-Indo-Europeanising of English (I). // *Hitotsubashi Journal of Arts and Sciences*, Vol. 33, The Hitotsubashi Academy, 27 – 58.

Greenberg 1960: Greenberg, J. A quantitative approach to the morphological typology of language. // *International Journal of American Linguistics* Vol. 26 No. 3, 178 – 194. <https://archive.org/details/sim_international-journal-of-american-linguistics_1960-07_26_3/page/178/mode/2up> (accessed July 2022)

Haspelmath, Michaelis 2017: Haspelmath, M., Michaelis, S. M. Analytic and synthetic: Typological change in varieties of European languages. // Buchstaller, I., Siebenhaar, B. eds., *Language Variation: European Perspectives VI: Selected Papers from the Eighth International Conference on Language Variation in Europe (ICLaVEe 8)*, Amsterdam, The Netherlands: John Benjamins Publishing Co., 3 – 22.

Hinrichs 2004: Hinrichs, U. Wann, warum und wie wurde das Bulgarische zu einer analytischen Sprache? – Ein Beitrag zur Entmystifizierung der Slavistik. // *Zeitschrift für Slawistik*, Vol. 49, No. 4, 381 – 419. <https://doi.org/10.1524/slaw.2004.49.4.381> (accessed July 2022)

Hollmann 2009: Hollmann, W. Grammatical change. // Culpeper, J., Katamba, F., Kerswill, P., McEnery, T. (eds.), *English language: description, variation and context*, Basingstoke: Palgrave, 314 – 333.

Horsch 2021: Horsch, J. Typological profiling of English, Spanish, German and Slovak: A corpus-based approach. // *Journal of Linguistics/Jazykovedný casopis* Vol. 72 No. 2, 342 – 352. <https://intapi.sciendo.com/pdf/10.2478/jazcas-2021-0032> (accessed July 2022)

Kortmann, Szmrecsányi 2009: Kortmann, B., Szmrecsányi, B. World Englishes between simplification and complexification. // Hoffmann, T., Siebers, L. (eds.), *World Englishes – Problems, Properties and Prospects: Selected papers from the 13th IAWE conference*. John Publishing Company, 263 – 286

Ledgeway 2011: Ledgeway, A. Syntactic and morphosyntactic typology and change. // Maiden, M., Smith, J. C., Ledgeway, A. (eds.), *The Cambridge History of the Romance Languages: Volume I: Structures, Cambridge*, UK: Cambridge University Press, 382 – 471.

Levshina 2020: Levshina, N. How tight is your language? A semantic typology based on Mutual Information. // Evang, K., Kallmeyer, L., Ehren, R., Petitjean, S., Seyffarth, E. Seddah, D. (eds.), *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*. Düsseldorf, Germany: Association for Computational Linguistics, 70 – 78

McWhorter 2001: McWhorter, J. What people ask David Gil and why: Rejoinder to the replies. // *Linguistic Typology* Vol. 5, 388 – 413.

Osenova 2010: Osenova, P. Bulgarian. // *Revue belge de philologie et d'histoire*, Vol. 88, No. 3, 643 – 668. <https://doi.org/10.3406/rbph.2010.7798> (accessed July 2022)

Schlegel 1818: Schlegel, A. W. *Observations sur la langue et la littérature provençales*, Paris, France: Librairie grecque-latine-allemande. <https://gallica.bnf.fr/ark:/12148/bpt6k63239579/f24.item.texteImage#.> (accessed July 2022)

Schmid 1994: Schmid, H. *Improvements in Part-of-Speech Tagging with an Application to German. Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf> (accessed Sept. 2023)

Schmid 1995: Schmid, H. *Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf> (accessed Sept. 2023)

Schwegler 1990: Schwegler, A. *Analyticity and Syntheticity: A Diachronic Perspective with Special Reference to Romance Languages*, Berlin/New York: Mouton de Gruyter

Siemund 2004: Siemund, P. Analytische und synthetische Tendenzen in der Entwicklung des Englischen. // Hinrichs, U. (ed.), *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp.* Harassowitz

Simov, Osenova and Slavcheva 2004: Simov, K., Osenova, P., Slavcheva, M. BTB-TR03: BulTreeBank Morphosyntactic Tagset* BTB-TS version 2.0. Linguistic Modelling Laboratory, Bulgarian Academy of Sciences. <http://bultreebank.org/wp-content/uploads/2017/06/BTB-TR03.pdf> (accessed Sept. 2023)

Stankov 2024: Stankov, N. Development trends in the morphological-typological profile of German. [Manuscript submitted for publication] // *Studia philologica*, Vol. 43 No. 1

Szmrecsányi 2009: Szmrecsányi, B. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. // *Language Variation and Change*, Vol. 21, Cambridge University Press, 319 – 353.

Szmrecsányi, Kortmann 2011: Szmrecsányi B., Kortmann, B. Typological profiling: Learner Englishes versus indigenized L2 varieties of English. // Mukherjee, J., Hundt, M. (eds.), *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*, Amsterdam/Philadelphia, The Netherlands/USA: John Benjamins Publishing Co., 167 – 188.

Szmrecsányi 2016: Szmrecsányi, B. An analytic-synthetic spiral in the history of English. // van Gelderen, E. (ed.), *Cyclical Change Continued*, 93 – 112. <https://doi.org/10.1075/la.227.04szm> (accessed Sept. 2023)

Tristram 2009: Tristram, H. Wie weit sind die inselkeltischen Sprachen (und das Englische) analytisiert? // Hinrichs, U., Reiter, N., Tornow, S. (eds.), *Eurolinguistik: Entwicklung und Perspektiven*, Wiesbaden, Germany: Harrassowitz, 255 – 280.

Uktamovna 2023: Uktamovna, K. Comparative typology of the English and Uzbek languages. // *International Journal of Literature and Languages*, Vol. 3, No. 6, 44 – 48.