

TYPOLOGICAL PROFILING OF ENGLISH, SPANISH, GERMAN AND SLOVAK: A CORPUS-BASED APPROACH

JAKOB HORSCH

Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany

HORSCH, Jakob: Typological profiling of English, Spanish, German and Slovak: A corpus-based approach. *Journal of Linguistics*, 2021, Vol. 72, No 2, pp. 342 – 352.

Abstract: Inspired by earlier work on typological profiling of English by Benedikt Szmrecsányi and Bernd Kortmann ([1], [2], [3]), this paper investigates the typological profiles of English, Spanish, German, and Slovak, applying Szmrecsányi and Kortmann’s methodology of calculating a SYNTHETICITY INDEX and an ANALYTICITY INDEX based on 1,000-word corpus samples. The results show that Szmrecsányi and Kortmann’s methodology is replicable, and confirm claims in the literature about degrees of analyticity and syntheticity of these languages. Instead of a simple analytic-synthetic continuum, Szmrecsányi and Kortmann’s “typological space” [3] is used to visualize results, showing that languages can be both synthetic *and* analytic to varying degrees.

Keywords: typological profiling, syntheticity index, analyticity index, typological space, English, German, Spanish, Slovak, corpus samples

1 INTRODUCTION

In morphological typology, the terms *synthetic* and *analytic* are widely used to describe languages based on their morphosyntactic properties. Accordingly, languages are characterized “as rather analytic [...] or as rather synthetic” [1]. English, for example, has been referred to as “analytic to a very high degree” [4], and Slovak is considered a synthetic, or inflecting language: It is described as “výrazne flektívna” (‘significantly inflecting’) [5] or “Slovenčina je prevažne flektívny jazyk” (‘predominantly inflecting’) [6]. In-between the analytic and synthetic extremes are languages like Spanish, which has “retained a large number of synthetic verb forms while undergoing a radical change towards analyticity in the domain of nouns and adjectives”, and German, where “the verb phrase [...] is one of structural and lexical analyticity [...] combined with a fairly high degree of syntheticity in the maximally governing finite verb” [4].

But just *how* synthetic or analytic are languages? Statements like ‘analytic to a very high degree’, ‘predominantly inflecting’, ‘radical change towards analyticity’, and ‘fairly high degree of syntheticity’ remain somewhat vague. Addressing this issue, the main objective of the present study is to determine and compare the typological (i.e., morphosyntactic) profiles of four Indo-European

languages from across the spectrum that ranges from analytic to synthetic: English, Spanish, German, and Slovak. Following the methodology of Szmrecsányi and Kortmann [3], degrees of analyticity and syntheticity will be calculated based on random samples drawn from corpora.

Before proceeding, a brief¹ overview of the history of the terms *synthetic* and *analytic* will be provided – after all, they “have a long and venerable tradition in linguistics” [1]. They were coined in the early 19th century by August Wilhelm von Schlegel² [7], whose “simple binary classification” [8] was refined by Sapir [9] into a “scalar concept of syntheticity” [8]. Building on Sapir’s work, Greenberg devised a mathematical formula for calculating a “synthetic index”, which he defined as “the ratio M/W where M equals morpheme and W equals word”, such that “[a]nalytic languages will give low results on this index, synthetic higher” [10].

Although Greenberg’s formula is very appealing because of its simplicity, reality is more complicated, which is reflected in the fact that the terms ‘analytic(ity)’ and ‘synthetic(ity)’ are not used consistently in the literature. Szmrecsányi, for example, points out “terminological confusion” [1], and Schwegler laments a “vagueness of terms” [8] in this context. Therefore, precise definitions are in order at this point.

The approach to analyticity and syntheticity adopted in the present study is that of Szmrecsányi and Kortmann, who have done some groundbreaking work. Inspired by Greenberg’s syntheticity index, Szmrecsányi and Kortmann defined “overt *grammatical syntheticity*” as “the text frequency of bound grammatical markers” [3] (emphasis in the source), and “overt *grammatical analyticity*” as “the text frequency of free grammatical markers” [3] (emphasis in the source), adding a dimension that had largely been ignored previously.³ Thus, languages could be profiled in more detail by providing measurements of syntheticity *and* analyticity. Whereas Greenberg’s synthetic index was used to describe degrees of syntheticity *versus* analyticity, with Szmrecsányi and Kortmann’s approach it was now possible to describe degrees of syntheticity *and* analyticity.

As the terms *grammatical syntheticity* and *grammatical analyticity* indicate, Szmrecsányi and Kortmann [3] focus on the marking of grammatical information, disregarding lexical processes such as derivation and compounding. Furthermore, Szmrecsányi and Kortmann’s approach is only concerned with *overt* grammatical marking, that is, the *presence* of morphemes, ignoring phenomena such as null

¹ The reader is referred to chapter 1 in Schwegler [8] for a detailed overview (as pointed out by Szmrecsányi [1]).

² Schlegel originally used the terms ‘synthetic’ and ‘analytic’ to distinguish between the evolutionary stages of inflectional languages, as noted by Askedal [4].

³ Although, as Szmrecsányi notes [1], the need for an analyticity index was noted as early as the 1980s by Kasevič and Jachontov [11].

marking, or zero morphemes.⁴ They do, however, take into account “allomorphs including ablaut phenomena [...] and other nonregular, yet clearly bound grammatical markers” such as suppletion [1]. Szmrecsányi and Kortmann’s precise definition⁵ of grammatical analyticity and grammatical syntheticity is thus as follows:

“[F]ormal grammatical analyticity [is defined] as comprising all those coding strategies where grammatical information is conveyed by free grammatical markers, which we in turn define as function words that have no independent lexical meaning. Conversely, we take *formal grammatical syntheticity* to comprise all those coding strategies where grammatical information is signaled by bound grammatical markers.” [3] (emphasis in the source)

Based on this definition, Szmrecsányi and Kortmann devised the following two formulas, which will be applied in the present investigation:

(1) “The *analyticity index*: the ratio of the number of free grammatical markers in a sample (F) to the total number of words in the sample (W), normalized to a sample size of 1,000 tokens. Hence: ANALYTICITY INDEX = $f/w \times 1,000$.” [3] (emphasis in the source)

(2) “The *syntheticity index*: the ratio of the number of words in a sample that bear a bound grammatical marker (B) to the total number of words in the sample (W), normalized to a sample size of 1,000 tokens. Hence: SYNTHETICITY INDEX = $b/w \times 1,000$.” [3] (emphasis in the source)

Szmrecsányi and Kortmann’s work focused on English and its varieties, e.g., intra-lingual variation in English across different registers [1], comparing Learner Englishes to L2 varieties of English [3], and tracing the diachronic evolution of English [2]. However, what is of much greater interest in the context of cross-linguistic typology is what appears to have been more of a by-product of one of their studies. To “investigate the issue of substrate effects” on Learner Englishes [3], Szmrecsányi and Kortmann included the following six European languages: Bulgarian, Czech, French, German, Italian, and Russian. Table 1 provides an overview of their results: the analyticity and syntheticity indices of these languages, based on 1,000-word corpus samples.

The numbers correspond to what most linguists would intuitively predict. Low syntheticity index (SI) scores were determined for English (SI: 197) and French (SI:

⁴ Incidentally, most grammars that were consulted in the context of the present study explicitly reject the concept of a zero morpheme (e.g. the approaches of the Real Academia Española’s [12] and the *Duden*, a grammar of German [13]). However, some do not (e.g. Dvonč et al. [5] and Oravec [6]).

⁵ Note that Szmrecsányi and Kortmann’s definitions are “strictly *formal* [...] and not semantic” [1] in nature. Thus, the multiple meanings of portmanteau morphs are disregarded in calculating the syntheticity index.

153), which are analytic languages (as noted by, e.g., Oravec [6]). Conversely, Czech (SI: 683) and Russian (SI: 670), two synthetic languages, have the highest SI scores and the lowest analyticity index (AI) scores (Czech AI: 334; Russian AI: 300). English (AI: 427) and French (AI: 439) score high in this regard, and German (SI: 301, AI: 436), Bulgarian (SI: 394, AI: 372) and Italian (SI: 250, AI: 458) cover the middle ground.

<u>Language</u>	<u>Analyticity index</u>	<u>Syntheticity index</u>
(British) English ⁶	427	197
Bulgarian	372	394
Czech	334	683
French	439	153
German	436	301
Italian	458	250
Russian	300	670

Tab. 1. Analyticity and syntheticity index values of various European languages, data from Szmrecsányi and Kortmann [3]

Note, however, that these results are based on “comparatively small corpora” consisting of “approx[imately] 10,000 words of running text each [...] sampling newspaper prose” [3]. Furthermore, Szmrecsányi and Kortmann had part-of-speech (POS) annotation carried out by a different person for each language, “typically by native speakers” [3]. While there are good reasons for recruiting native speakers, who can be expected to have sound knowledge of their L1 languages, Szmrecsányi and Kortmann’s approach comes with the disadvantage that there might have been inconsistencies in coding, affecting the comparability of their results.

The present study, apart from shifting Szmrecsányi and Kortmann’s focus on synchronic and diachronic *intra*-linguistic variation in English to *inter*-linguistic variation, seeks to address these issues by using random samples from much larger corpora and having data annotation carried out by the same researcher. The languages chosen were English, Spanish, German, and Slovak, based on the following considerations: (1) To test the feasibility of applying Szmrecsányi and Kortmann’s methodology to further languages (i.e., Spanish and Slovak); (2) to see whether Szmrecsányi and Kortmann’s study could be replicated with regard to English and German; (3) because English and Slovak fall on the opposite ends of the analytic-synthetic continuum, with Spanish and German covering the middle ground; and (4) because the author is fluent in all four of these languages, so they could be coded with a high degree of consistency.

⁶ Szmrecsányi and Kortmann actually investigated three different registers of the British National Corpus (BNC), including conversation, university essays, and school essays. The value provided here is that of university essays [3], as this register appears to be the most comparable to the newspaper prose that was sampled for the other European languages.

The main objective of the present study is, thus, to calculate, following Szmrecsányi and Kortmann’s methodology, the analyticity and syntheticity indices of English, Spanish, German, and Slovak. The data that will serve as a basis for these calculations consists of 1,000-word random samples extracted from four *Sketch Engine* corpora [14], [15]. Since these corpora were compiled by the same researchers, there should be a high degree of comparability between languages.

The expected results are as follows: Based on Szmrecsányi and Kortmann’s findings, English should score high in terms of analyticity and low in syntheticity. Slovak, a synthetic language, should have scores similar to its close relative Czech⁷, for which Szmrecsányi and Kortmann determined a low analyticity index and a high syntheticity index (see table 1). The analyticity and syntheticity indices of German, for which Szmrecsányi and Kortmann determined an AI of 436 and an SI of 301, as well as Spanish, for which no such indices have been calculated so far, should fall somewhere in-between English and Slovak.

2 DATA AND METHODOLOGY

The main objective of this study, as outlined in section 1, is to determine the analyticity and syntheticity indices of English, Spanish, German, and Slovak. The calculations are based on data samples from the Slovak Web 2011 (skTenTen11), the Spanish Web 2018 (esTenTen18), the German Web 2013 (deTenTen13), and the English Web 2015 (enTenTen15) corpora, which were queried using Sketch Engine’s online interface⁸ in July and December 2020; these corpora were chosen for reasons of comparability.⁹ CQL queries were used to extract all words from each corpus, excluding punctuation and other symbols.¹⁰ Subsequently, a 1,000-word sample was drawn by means of Sketch Engine’s “get a random sample” function. These samples were downloaded as CSV files and annotated in *LibreOffice Calc* [17], and then saved as ODS files.¹¹ Each token from the random sample was subsequently annotated for the following variables:

- FUNCTION WORD (levels: TRUE, FALSE)
- NUMBER OF BOUND GRAMMATICAL MARKERS (levels: 0, 1, 2...)

⁷ Actually, the two languages are so closely related to each other that they are generally considered to form a “dialect continuum” [16].

⁸ Access to Sketch Engine (<https://app.Sketch Engine.eu>) was generously made available to the Catholic University of Eichstätt-Ingolstadt through the ELEXIS Program (<https://elexis.is>).

⁹ As one reviewer noted, the Slovak National Corpus could have been used to achieve a higher degree of representativeness. However, in the interest of obtaining comparable samples, it was decided to use skTenTen11, which is more similar to the other corpora in that it contains texts from the web only and was compiled by the same researchers.

¹⁰ The full CQL expressions are contained in the files at the research project’s OSF repository (see below).

¹¹ All files – including the original CSVs downloaded from Sketch Engine and the manually coded ODS files – are available at the following OSF repository: <https://osf.io/9w3u5/>.

Annotation and morphological segmentation to determine the number of bound grammatical markers was carried out based on the following standard grammars:

- English: *A Comprehensive Grammar of the English Language* [18],
- German: *Duden: Die Grammatik* [13],
- Spanish: *Nueva gramática de la lengua española* [12],
- Slovak: *Morfológia slovenského jazyka* [5]; *Morfológia spisovnej slovenčiny* [6].

The coding of the variable FUNCTION WORD, which involved checking whether the word token was “synsemantic”, i.e., with “no independent lexical meaning” [1], was greatly facilitated by POS annotation in the corpora. Nevertheless, each token was manually checked, as corpora have been known to contain erroneous tags. Each word token with a POS tag corresponding to closed word classes loaded on to the analyticity index of the respective language. This included prepositions, pronouns¹², determiners, conjunctions, modal/auxiliary verbs, negators, primary verbs in auxiliary function (English), the infinitive marker *to* (English), and particles.

Next, the number of bound grammatical markers (i.e., affixes) was counted. To do so, each word token was segmented using the paradigmatic substitution test to determine whether a morpheme carried meaning. Thus, the Slovak word *pripravila* ‘she prepared’ would be segmented into three morphemes; a stem (*pripravi-*) and two grammatical affixes that indicate past tense (*-l-*) and gender (*-a*). Regarding English, this was a rather simple undertaking, since no more than one grammatical suffix can attach to a word at a time. In the case of Spanish, German, and Slovak, however, matters were more complex:

- In German, there are certain noun inflection classes whose dative plural endings can be segmented into two separate morphemes, e.g., *den Tag-e_{PL}-n_{DAT}* [19]. Also, certain past tense forms of verbs have two segmentable affixes, e.g., *such-te_{PST}-st_{2SG}* [19]. Circumfixes, e.g., *ge_{PST}-sag-t_{PST}*, however, were only counted as one morpheme, as were combinations of *umlaut* with suffixation, e.g., *der Turm_{SG} → die Türm-e_{PL}*;
- In Romance languages including Spanish, “number and gender marking on nouns and adjectives is [...] typically suffixal” [20], so that these word classes can carry up to two distinct suffixes, e.g., *niñ-o_M-s_{PL}*.¹³ Similarly, inflected Spanish verbs can carry up to two suffixes [12], e.g. *cantá-ba_{IMP}-mos_{1PL}*.¹⁴

¹² Although note that some Slovak grammars exclude pronouns from the group of synsemantic words [6].

¹³ Concerning the approach to the grammatical gender morpheme in Spanish in the present paper, cf. [12].

¹⁴ Note that according to the *Nueva gramática de la lengua española*, inflected verbs in Spanish can actually be segmented into up to four components: root; thematic vowel; tense, aspect and mood marker; person and number marker [12]. In the present study, however, the thematic vowel was disregarded, as it does not carry any meaning [12], making its morpheme status (which in the present study is defined as ‘the smallest meaning-bearing unit’) disputable.

- Slovak superlative adjective forms can be segmented into three affixes (comparative, superlative, and gender), e.g. *naj_{SUP}-siln-ejš_{COMP}-ia_F*.

Thus, in Spanish, German, and Slovak, one word token could load on to the syntheticity index multiple times. Once the data was annotated, the “two Greenberg-inspired index values” [3] were calculated according to Szmrecsányi and Kortmann’s formulas (see section 1).

3 RESULTS AND DISCUSSION

After annotating the 1,000-token samples, the analyticity and syntheticity indices were determined for each language. Table 2 presents the results:

Language	Analyticity index	Syntheticity index
English	395	210
Spanish	423	410
German	458	517
Slovak	355	595

Tab. 2. Analyticity and syntheticity indices of English, Spanish, German, and Slovak

A first glance at table 2 confirms the expectations outlined in section 1. English scores low in syntheticity, with an SI of 210 – that is, out of 1,000 words, 210 bear a grammatical marker. In contrast, Slovak has the highest syntheticity score: the 1,000-word sample contained 595 inflectional morphemes. Spanish (SI: 410) and German (SI: 517) are in-between these two extremes. Slovak has the lowest AI score (AI: 355), i.e., only 355 function words were found in the 1,000-word sample. Interestingly, Spanish (AI: 423) and German (AI: 458) actually score *higher* than English (AI: 395) in this regard.

Regarding Szmrecsányi and Kortmann’s [3] results for English (AI: 427; SI: 197), the present study’s results (AI: 395; SI: 210) come very close. A Chi-squared test confirms that there is no statistically significant difference (AI: $\chi^2=1.25$, $df=1$, $p>0.5$; SI: $\chi^2=0.42$, $df=1$, $p>0.5$). Furthermore, Slovak (AI: 355; SI: 595) turned up similar scores as its close relative Czech, for which Szmrecsányi and Kortmann determined an AI of 334 and an SI of 683. This suggests that their method is indeed replicable.

However, German (AI: 458; SI: 595) deviates significantly from Szmrecsányi and Kortmann’s results (AI: 436; SI: 301) with regard to syntheticity, as a Chi-squared test confirms (AI: $\chi^2=0.54$, $df=1$, $0.5<p<0.1$; SI: $\chi^2=57.04$, $df=1$, $p<0.001$). One explanation might be that the texts from which the samples were drawn differed from each other. This is not entirely implausible – one of Szmrecsányi’s studies showed, for example, that “variability in analyticity and syntheticity is endemic, surprisingly so, even among closely related dialects and varieties of the same

language” [1]. For clarification, it would be necessary to compare the data to Szmrecsányi and Kortmann’s data set.

Figure 1 provides visualization of the results by means of Szmrecsányi and Kortmann’s “typological space”, where the y -axis “plots analyticity index scores while the x -axis indicates syntheticity index scores” [3]. Thus, a highly analytic language such as English will be in the top left corner (high analyticity, low syntheticity), whereas a synthetic language such as Slovak will be found in the lower right corner (low analyticity, high syntheticity). For comparison, the figure also contains Szmrecsányi and Kortmann’s [3] results for English, German, and Czech (as gray data points).

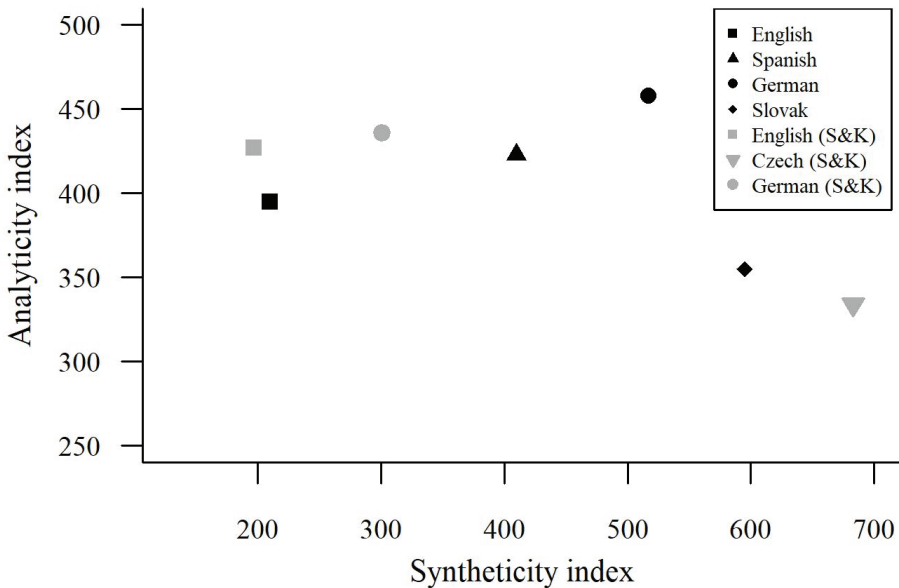


Fig. 1. Typological space: analyticity vs. syntheticity (in index points); index values from Szmrecsányi and Kortmann [3] in gray

The typological space facilitates comparison between languages. Figure 1 shows that in terms of syntheticity, the ‘in-between’ languages German and Spanish cover the middle ground between the ‘extremes’, English and Slovak. Another insight from the diagram is that analyticity and syntheticity are not necessarily exclusive categories or opposite poles of a one-dimensional continuum (as discussed in section 1): While English is indeed a “textbook example of a language that has developed from a synthetic language into an analytic one” [1], scoring high in analyticity and low in syntheticity, German and Spanish are not only more synthetic,

but also more analytic, a fact that is not easily appreciated without this kind of visualization.

4 CONCLUSION

The present study has demonstrated the feasibility of typologically profiling languages using Szmrecsányi and Kortmann's [3] methodology, that is, determining degrees of syntheticity and analyticity of languages based on naturalistic language data in the form of corpus samples. It was possible to confirm claims in the literature, such as English being "analytic to a very high degree" [4] and Slovak being "significantly inflecting" [5]. By calculating analyticity and syntheticity indices, such claims can now be substantiated with empirical evidence. It was also demonstrated that analyticity and syntheticity indices allow for a very precise comparison of languages.

It was furthermore shown that instead of an analytic-synthetic continuum, it is more appropriate to use a typological space consisting of two axes. This is because languages can be synthetic *and* analytic to varying degrees. It was also possible to replicate Szmrecsányi and Kortmann's [3] results regarding English and German, although German deviated considerably with regard to the syntheticity index. The present study has also highlighted the manifold possibilities in which corpus data can and should be employed in linguistics. Beyond exploring syntactic, lexical, and morphological phenomena, it can also be employed for the typological profiling of languages.

Finally, it must be noted that calculations of analyticity and syntheticity based on word/morpheme counts must always be taken with a grain of salt, because as usual, matters are more complex than they appear at first glance. As Schwegler notes, "many of the so-called analytic constructs [...] have a considerably tighter morphological cohesion (i.e., are more synthetic) than the label analytic suggests" [8]. In this context, he mentions that "[Old French] *je* 'I' has a morphosyntactic and semantic profile which in many ways parallels that of its [Latin] ancestor *ego*, but [...] in many ways differs fundamentally [...] by having entered into a tighter relation with the verb" [8]. Similarly, the English pronoun *I* is in a tighter relation with the verb than its counterpart *ja* in Slovak, which is a pro-drop language like Latin.

Likewise, one might argue that there are varying degrees of syntheticity. Consider the Slovak locative case, which is marked on nouns with a suffix. A noun with a locative suffix would, therefore, add to the syntheticity index. However, locative-case nouns never appear without a preposition in Slovak, having lead some to even speak of a hybrid synthetic-analytic strategy [6]. However, the present study's approach simply takes words and morphemes and drops them into one of two bins – synthetic or analytic – instead of taking into account the context (e.g.

personal pronouns or prepositions) within which they are found. To address concerns about varying degrees of analyticity (e.g. pronoun-verb relation in English) and syntheticity (e.g. locative case with obligatory prepositions in Slovak), future research will have to take into account the context of words and how strongly they are ‘attracted’ to each other.

The present small-scale investigation, which is to be understood as a pilot study, holds considerable potential for future research. Calculating the analyticity and syntheticity indices of other languages, it is possible to further test and corroborate (or refute, for that matter) claims in the literature, e.g., about typological relatedness. For example, Czech and Slovak have been described as “two closely related languages” [16] – the present study could not only confirm this claim, but also determine just *how* closely they are related. From a diachronic perspective, syntheticity and analyticity indices can help trace the morphosyntactic evolution of languages [2]. As Schwegler notes, “many long-term diachronic changes cannot be grasped appropriately without the notions of analyticity and syntheticity” [8]. Finally, one issue that arises from using the Sketch Engine corpora is that they are composed of texts from the internet, which raises questions about their representativeness. Future research should, therefore, be based on data from more balanced corpora.

ACKNOWLEDGEMENTS

Thanks go to the Sketch Engine support team for helping me refine my initially rather cumbersome CQL expressions. I also highly appreciate the comments of two anonymous reviewers who helped improve my manuscript.

References

- [1] Szmrecsányi, B. (2009). Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21(3), pages 1–35.
- [2] Szmrecsányi, B. (2012). Analyticity and syntheticity in the history of English. In T. Nevalainen and E. C. Traugott (eds.), *The Oxford Handbook of the History of English*, pages 654–665. Oxford: Oxford UP.
- [3] Szmrecsányi, B., and Kortmann, B. (2011). Typological profiling: learner Englishes versus indigenized L2 varieties of English. In J. Mukherjee and M. Hundt (eds.), *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, pages 167–187. Amsterdam: Benjamins.
- [4] Aske, J. O. (2001). Conceptions of typological change. In M. Haspelmath (ed.), *Language Typology and Language Universals: An International Handbook*, vol. 2, pages 1624–1639. Berlin: De Gruyter Mouton.
- [5] Dvonč, L., Horák, G., Miko, F., Mistrík, J., Oravec, J., Ružička, J., and Urbančok, M. (1966). *Morfológia slovenského jazyka* [‘Morphology of the Slovak language’]. (Ed.) J. Ružička. Bratislava: Slovenská akadémia Vied [‘Slovak Academy of Sciences’].

- [6] Oravec, J. (1980). *Morfológia spisovnej slovenčiny* [‘Morphology of written Slovak’]. Bratislava: Slovenské pedagogické nakladateľstvo [‘Slovak Publishing House of Educational Literature’].
- [7] Schlegel, A. W. von (1818). *Observations sur la langue et la littérature provençales* [‘Observations on the provençal language and literature’]. Paris: Librairie grecque-latino-allemande.
- [8] Schwegler, A. (1990). *Analyticity and syntheticity: a diachronic perspective with special reference to Romance languages* (Empirical Approaches to Language Typology 6). Berlin; New York: Mouton de Gruyter.
- [9] Sapir, E. (1921). *Language*. New York: Harcourt, Brace and Co.
- [10] Greenberg, J. H. (1960). A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3), pages 178–194.
- [11] V. Kasevič and S. E. Jachontov (eds.). (1982). *Квантитативная типология языков Азии и Африки* [‘A quantitative typology of Asian and African Languages’]. Leningrad: издательство ленинградского университета [‘Leningrad University Press’].
- [12] Real Academia Española. (2009). *Nueva gramática de la lengua española* [‘New grammar of the Spanish language’]. Vol. 1: *Morfología, Sintaxis*. Madrid: Espasa Libros, S. L. U.
- [13] A. Wöllstein (ed.). (2016). *Duden: Die Grammatik* [‘Duden: The Grammar’]. 9th updated and completely revised edition. Berlin: Dudenverlag.
- [14] Kilgarriff, A., Smrž, P., and Tugwell, D. (2004). The Sketch Engine. In G. Williams and S. Vessier (eds.), *Proceedings of the eleventh EURALEX international congress (EURALEX 2004)*, pages 105–116. Lorient: Université de Bretagne-Sud, France.
- [15] Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX 1*, pages 7–36.
- [16] Nábělková, M. (2016). The Czech-Slovak Communicative and Dialect Continuum: With and Without a Border. In T. Kamusella, N. Motoki and C. Gibson (eds.), *The Palgrave Handbook of Slavic Languages, Identities and Borders*, pages 140–184. London: Palgrave Macmillan.
- [17] The Document Foundation. (2021). LibreOffice Calc v. 7.1.2. Available at: <https://www.libreoffice.org/discover/calc/>.
- [18] Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. New York: Longman.
- [19] Gallmann, P. (2016). Die flektierbaren Wortarten [‘Inflective word classes’]. In *Duden: Die Grammatik* [‘Duden: The Grammar’], pages 149–385. 9th updated and revised edition. Berlin: Dudenverlag.
- [20] Ledgeway, A. (2017). Syntheticity and Analyticity. In A. Dufter and E. Stark (eds.), *Manual of Romance Morphosyntax and Syntax*, pages 839–886. Berlin: De Gruyter Mouton.