

Some Typological Characteristics of Czech and English and Other European Languages

Milan Malinovský

Abstract

Numerous attempts have been made to compare different types of languages from the viewpoint of their quantitative characteristics. We present a piece of research in five Indo-European languages focused on length of sentences and words, not drawing any general conclusions but bringing some partial findings about these languages' general measurable features.

1. Introduction

This article deals with the differences between quantitative characteristics of individual languages¹ comparing in particular the length of sentences and words in contemporary Czech and English. Several texts in German, French and Polish have also been included for comparison in order to illustrate the case in the three major European language families, i.e. the Germanic, Romance and Slavonic languages; in other words, to highlight certain structural differences between inflectional and analytic languages.

The aim of the study is to examine and possibly prove that in analytic languages sentences contain more words.

As well as the length of sentences, i.e. the number of words that they consist of, the focus is also on words (and the number of their constituent syllables). Detailed summaries of our research are given in Part Four.

All the texts examined are contemporary; always, a pair (triad, ...) of original + translation are used to illustrate the point. In every case, only published translations were used, in order to avoid the danger of tentative subjectivity of one's own translation while judging the individual phenomena.

In spite of the fact that we have used different genres (cf. Literature) and researched a great number of texts², we are fully aware of the fact that our findings and

conclusions are only preliminary and should be interpreted with caution. For real quantitative conclusion to be drawn about the relationships between pairs of texts in various languages (and between different types of language), a more extensive study, utilizing a greater corpus of material, would be required.

The majority of texts used are drawn from the field of philosophy, as we believe they are fairly appropriate for research of this kind because of their (neither very concise nor excessively verbose) style.³ The other texts come from history, political science, ecology, sociology, and so forth.

A general comment on translation: if it is true that a (good) translator is in fact a co-author of a text (and therefore not only a transposer) then it is not strictly possible to make an entirely objective comparison about the lengths of equivalent sentences in two languages (however representative they may be).

2. Method of research

The research proceeded as follows: owing to the different formats of the individual language sources, we transposed the results of calculations to 'standard' pages (30 lines times 60 keystrokes) for the comparison be utterly precise. Data was acquired by comparing these standard pages for the length of the text in the first and second (third, ...) language.

2.1. Disparateness of languages: sentences

As might be expected, individual texts in the languages examined were of different lengths. Because of this, it was not technically possible to compare the lengths of parallel sentences reciprocally. Nonetheless, some comparisons of individual sentences were made; this was the case where the total number of sentences in a given text was not the same for the two (or more) languages being compared. One of the main reasons for such a disparity was the fact that in some instances a compound sentence in the original text (common in French⁴, for example) was broken up into two (or three) sentences in translation.

As mentioned above, the majority of the source material was taken from philosophical texts. Now, let us examine the composition of a text in its individual translations in the following compound sentence (Patočka, 1972, p. 14; its translations are given in Literature):

(French): *Husserl ne l'a pas d'ailleurs développée complètement; l'ouvrage est resté inachevé; la solution qui lui servait de fil conducteur se laisse définir le plus clairement à l'aide d'une conférence prononcée par Husserl à Vienne en 1935; là, il brosse l'image d'un monde naturel, résultat d'achèvements (Leistungen) en commun des subjectivités dont ce monde constitue précisément un trait d'union sans substantialité propre; c'est le retournement complet de la vue physicaliste, il est vrai, un reversement plus que copernicien - au lieu d'îlots de subjectivité contournés d'une mer infinie de structures physicalistes, schemas abstraits dépourvus de*

qualités et de vie, on a une mer d'intersubjectivité dite transcendente qui entoure des unités objectives constituées en unités de sens par le travail en commun des consciences communiquantes.

(Czech): Husserl ji také nikdy úplně nerozvinul, jeho dílo zůstalo nedokončeno. Řešení, které mu bylo vodítkem, lze nejzřetelněji určit z přednášky, kterou proslovil ve Vídni roku 1935. Tam načrtává obraz přirozeného světa jako výsledku společných výkonů (Leistungen) subjektivit, jimž tento svět není ničím jiným než pojítkem bez substanciality ve vlastním smyslu. Je to vskutku úplně převrácení fyzikalistického pohledu, obrat víc než koperníkovský: místo ostrůvků subjektivity obklopených nekonečným mořem fyzikalistických struktur, abstraktních schémat zbaivených kvalit i života, je zde moře intersubjektivit, zvané transcendentální, jež obklopuje objektivní jednotky konstituované společnou prací komunikujících vědomí v jednotky smyslu.

(English): (The first sentence of the compound sentence stayed untranslated.) We can best glean the solution that guided him from the lecture he delivered in Vienna in 1935. There he sketches the image of a "natural" world conceived as a product of the common achievements (Leistungen) of subjectivities for whom this world is nothing but a common link devoid of any genuine substantiality. It is, in truth, a complete inversion of the physicalist view, a more than Copernican reversal. In place of islets of subjectivity surrounded by an infinite sea of physicalist structures, of abstract schemata stripped of qualities and of life, we now have a sea of allegedly transcendental intersubjectivity surrounding objective unities constituted as unities of meaning by the common efforts of communicating consciousnesses.

(Polish): Zresztą Husserl nie rozwinął jej w pełni; dzieło pozostało niedokończzone. Rozwinięcie, które posłużyło mu za nic przewodnią, można najwyraźniej określić w oparciu o wykład wygłoszony przez Husserla w Wiedniu w r. 1935. Maluje on w nim obraz świata naturalnego, wynik wspólnych dokonań (Leistungen) subiektywności, w których ów świat konstytuuje właśnie linię łączącą bez właściwej istotowości; jest to całkowite odwrócenie się od fizykalnego sposobu widzenia, przewrót większy niż kopernikański: zamiast wysepek subiektywności opasanych nieskończonym morzem struktur fizykalistycznych, abstrakcyjnych schematów pozbawionych jakości i życia, mamy morze intersubiektywności nazywanej transcendentálną, które otacza obiektywne jedności ukonstytuowane w jednościach sensu wspólną pracą komunikujących się świadomości.

(German): Hinzu kommt, daß Husserl sie nicht vollständig entfaltet hat; das Vorhaben blieb unvollendet. Im Wiener Vortrag von 1935 wird sein Leitgedanke am deutlichsten. Husserl zeichnet die natürliche Welt hier als Resultat von gemeinschaftlich erbrachten Leistungen der Subjekte, für welche diese Welt ein vereinigendes Element ohne eigene Substantialität darstellt. Er vollzieht eine totale Umkehrung der physikalistischen Sichtweise, eine mehr als kopernikanische Revolution: Statt Inseln der Subjektivität, die von einem unendlich weiten Meer physikalistischer Strukturen und abstrakter, lebloser und aller Qualität beraubter Schemata umgeben sind, haben wir hier das Meer einer als transzendental bezeichneten Intersubjektivität, welches objektive Einheiten, Sinneinheiten umschließt, die in der gemeinsamen Anstrengung kommunizierender Bewußtseine konstituiert wurden.

If a semicolon is considered a symbol of apposition within a compound sentence, it appears that the compound sentence of the French original (one of the most developed examples in our excerpts, consisting of 121 word units) was, when translated into four languages, translated as three or four separate sentences. This phenomenon was observed more than once in this text. Remarkably, both translators into Germanic and Slavonic languages chose to do this the same way (there are three sentences in the Polish translation of similar composition to the Czech translation). Moreover, the longest compound sentence was found in this very text (from the complete material researched): - the French version comprised 135 words, while the longest English compound sentence had 129 words. Incidence of such long sentences was more scarce in the other languages examined.

Generally speaking, the greater number of words observed in the sentences of analytic languages may be explained by the fact that in these languages the grammatical values are expressed by auxiliary words (referring, for example, to tense, person, and the category of determination). Similarly, while the Slavonic languages express syntactic relations of nouns prototypically with the help of case endings, the Germanic and Romance languages apply prepositions.

2.2. Disparateness in languages: words

In analyzing words, we considered mainly the graphic form of the word. English is a language abundant in one-syllable words. Let us illustrate this by using an excerpt of the English translation of *Karteziánství a fenomenologie* (Patočka, 1989, p. 307): *This world by means of which is not a world of beings, rather, it is I, a being that understands beings, that is open for them; and the core of this world which I am is a ...*, in which thirty words out of thirty-seven are one-syllabic.

It is of interest whether the phonetic realization in this case coincides with the written form, or, if and when only one out of two written syllables is pronounced (in Slavonic languages no differences of this kind are found). In order to note such differences, we considered both possibilities. This is restricted mainly to English; as the research has revealed, when counting the graphic syllables, English is rather consistent with Czech, but in its phonetic realization an English syllable unit is more likely to be a word (in comparison with Czech, it is multi-lettered, cf. *since, those, scratched*, etc.), see Part Four.

3. Conclusions

Contrary to our expectations, not many long words were found in German; in English, Czech, French, Polish and German, words longer than seven syllables are rare. The one and only exception was a 10-syllable German adjective *informationstheoretischer*, in Czech there was a 9-syllable genitive form of *psychofyziologického* and 11-syllabic locative form *čtyřiaadvacetikilometrovém*. For German (and English) conso-

nantal clusters are fairly common: in spite of the fact that the words look long, they are not very extensive syllablewise.

The excerpts indicate that presuppositions following from the typological characteristics of individual languages would be confirmed should more extensive language material be examined. In Czech, which is inflectional, there are fewer words on average than in English, although words are longer. The total length of a text (measured in syllables) is longer in Czech which means that a Czech word form is often longer than the corresponding English analytic form, and that Czech derived words are longer than English motivated word-forms. Czech can often easily do without auxiliaries. The length of the English written text is produced by specific features of English spelling. It should be noted that in philosophical texts, sentences are a little longer on average than in texts of other language styles. As mentioned, our collection of texts was too small for us to be able to draw any general conclusions from these partial findings.

As for the other languages, our findings suggest that in the main indices, French is in recognizable accordance with English as is Polish with Czech – again, in agreement with prediction based on typology. In some respects, German is closer to the Slavonic languages (in terms of word length), while in others it is closer to English and French (the length of sentences).

If research aimed in such a direction continues, the results may be very interesting for both general and comparative linguistics.

4. Statistical indices

The following are the numbers of pages (= standard pages), sentences, words, and syllables of the excerpted texts (in Czech and in English the extent of the texts was the same; other languages sets were restricted).

	Number of pages		sentences		words***	
Cz	264.6	81.7*	2,858**	693*	63,325	19,878*
En	278.8	93.5*	2,716**	714*	76,546	26,952*
Ge	101.4	101.4*	695	695*	24,115	24,115*
Fr	26.9	26.9*	195	195*	7,088	7,088*
Po	25.4	25.4*	202	202*	5,944	5,944*
Total	697.1	328.9*	6,666	2,499*	177,018	83,977*

Table 1. Total number of pages, sentences, and words in the texts of five languages (Czech, English, German, French, Polish).

* Part appertaining to philosophical text.

** Although the same number of texts was examined in both Czech and English, the significant difference in the number of sentences is due mainly to one of the sources in which the Czech translator divided the compound sentence of the original into more than one sentence.

*** The number of words was taken from the average number of words on a page (every fifth page).

As far as comparison of the number of words in one philosophical text (Patočka, 1972, see its translations in Literature) is concerned, the sequence is as follows: English (highest number of words), French, German, Polish, Czech (lowest).

	words on page		Average number of words in sentence		syllables in word	
Cz	239.5	246.4*	23.7	28.4*	2.4	2.4*
En	276.6	287.6*	29.5	36.4*	1.9 ^I	1.9* ^I
					1.7 ^{II}	-
Ge	-	241.2*	-	33.6*	-	2.1*
Fr	-	263.7*	-	36.3*	-	2.0*
Po	-	236.2*	-	29.7*	-	2.4*

Table 2. Average number of words on page, words in sentence, and syllables in each word.

* Philosophical text.

^I Graphic form (cf. par. 2.2.).

^{II} Phonetic realization (ibid.).

	Average number of n-syllabic words (in %)			
	Cz		En	
1 syllable	10.17	12.55 *	33.21 ^{II}	30.26 * ^I
2 syllables	24.56	23.80 *	23.86 ^{II}	18.28 * ^I
3 syllables	27.28	25.50 *	23.46 ^{II}	19.55 * ^I
4 syllables	21.10	17.47 *	15.14 ^{II}	16.87 * ^I
5 syllables	11.39	13.15 *	3.26 ^{II}	8.64 * ^I
6 syllables	3.54	5.11 *	0.76 ^{II}	4.08 * ^I
7 syllables	1.37	2.40 *	0.17 ^{II}	2.28 * ^I

Table 3. Percentage of occurrence of words according to number of syllables: Czech, English.

* Philosophical text.

^I Graphic form.

^{II} Phonetic realization.

	Occurance of words (in %)		
	Ge	Fr	Po
1 syllable	22.31 *	24.01 *	12.23 *
2 syllables	23.38 *	21.30 *	25.16 *
3 syllables	22.15 *	22.90 *	26.44 *
4 syllables	16.32 *	16.70 *	16.69 *
5 syllables	9.68 *	5.40 *	12.61 *
6 syllables	4.71 *	4.91 *	4.85 *
7 syllables	1.42 *	4.43 *	1.95 *

Table 4. Percentage of occurrence of words according to number of syllables: German, French, Polish.

* Philosophical text.

4.1. Ratio of the number of words in a sentence of philosophical texts

Sequence of languages: English, Czech, German (Patočka, 1976). In the given text, the English sentence has an average of 28.24 % more words than a Czech one, and 8.20 % more than a German one, that is, a German sentence has an average of 15.63 % more than a Czech one.

4.1.1. English, Czech, German, French, Polish (Patočka, 1972)

In the text provided, the English sentence has an average of 19.23 % more words than a Czech one, 6.37 % more than a German one, 9.11 % fewer than a French one, and 11.36 % more than a Polish one.

And vice versa: the Czech sentence has an average of 10.97 % fewer words than a German one, 23.77 % fewer than a French one, and 6.61 % more than a Polish one.

NOTES

¹ Within the field of quantitative linguistics, L. Uhlířová (1995) applies quantitative analysis to examine word length in Czech texts. Her findings lead to reasonable generalizations, modelled in a mathematically exact way by a general type of statistical distribution. Uhlířová claims that the distribution of word length in text is not only text-specific, text-distinctive, or even "random", but that it is subject to a more general probabilistic law. Her minute research brings evidence that the probabilistic laws found out for Czech are in full accordance with analogical probabilistic laws of the same generality which hold for other languages as well.

² The author expresses his thanks to Eva dos Reis for collating the statistical data.

³ Nacherová (1994) in her study points out that "a linguistic dimension of philosophical problems has its substantiation since it is the very linguistic aspect that is, as one of manifold aspects of formulation of philosophical problems, the most significant part within the process of formulation."

⁴ For completeness, let us mention that Patočka's study on "Edmund Husserl's Philosophy of the Crisis of the Sciences and His Conception of a Phenomenology of the 'Life-World'" was originally written (and later presented in Warsaw in 1971) in French.

REFERENCES

- NACHEROVÁ, S., 1994. Jazyk jako prostředek formulace filozofických problémů. Charles University College of Arts.
- SGALL, P., 1993. Typy jazyků a jejich základní vlastnosti. *Slovo a slovesnost*, 54, 271-277.
- SKALIČKA, V., 1941. Vývoj české deklinace. Praha.
- SKALIČKA, V., 1951. Typ češtiny. Praha.
- UHLÍŘOVÁ, L., 1995. On the Generality of Statistical Laws and Individuality of Texts. A Case of Syllables, Word Forms, their Length and Frequencies. *Journal of Quantitative Linguistics* 2, 238-247.

LITERATURE

- DIALOGUE, 1983. (Articles: Avoiding the Entropy Trap. Current Issues: Change and Continuity; Pluralism and Fundamentalism in the United States. Feminism, Family and Community. The New Conservatism.) The United States Information Agency Journal, 61, 16–33.
- GRAY, W., HOFSTADTER, R., OLSON, K. W., 1980. An Outline of American History. United States Information Agency, 1–52.
- GRAY, W., HOFSTADTER, R., OLSON, K. W., 1980. Nástin amerických dějin. Embassy of the United States of America. Praha, 1–52.
- PATOČKA, J., 1989. Cartesianism and Phenomenology. In: Jan Patočka, Philosophy and Selected Writings. The University of Chicago Press, Chicago & London, 285–326.
- PATOČKA, J., 1991. Cartesianismus und Phänomenologie. In: Jan Patočka, Die Bewegung der menschlichen Existenz. Phänomenologische Schriften II, Ernst Klett Verlag für Wissen und Bildung GmbH, Stuttgart, 360–414.
- PATOČKA, J., 1991. Die Philosophie der Krisis der Wissenschaften nach Edmund Husserl und sein Verständnis einer Phänomenologie der Lebenswelt. In: Jan Patočka, Die Bewegung der menschlichen Existenz. Phänomenologische Schriften II, Ernst Klett Verlag für Wissen und Bildung GmbH, Stuttgart, 310–329.
- PATOČKA, J., 1989. Edmund Husserl's Philosophy of the Crisis of the Sciences and His Conception of a Phenomenology of the "Life-World". In: Jan Patočka, Philosophy and Selected Writings. The University of Chicago Press, Chicago & London, 223–238.
- PATOČKA, J., 1980. Filosofie krize věd podle Edmunda Husserla a jeho koncepce fenomenologie "Světa našeho života". In: Přirozený svět a pohyb lidské existence. Samizdat edition, 3, Praha, 3.3.2–3.3.21.
- PATOČKA, J., 1987. Filozofia kryzysu nauki wedlug Edmunda Husserla i jego koncepcja fenomenologii "Świata życia". In: Jan Patočka, Świat naturalny i fenomenologia. Papieska Akademia Teologiczna, Wydział filozoficzny, Kraków, 140–157.
- PATOČKA, J., 1976. Karteziánství a fenomenologie. In: Samizdat edition, Praha, 1–38.
- PATOČKA, J., 1972. La philosophie de la crise des sciences d'après Edmond Husserl et sa conception d'une phénoménologie du "monde de la vie". In: Archiwum historii, filozofii i myśli społecznej, 18, Warszawa, 3–18.
- SAVICKÝ, P., HLAVÁČOVÁ, J., 2002. Measures of Word Commonness. Journal of Quantitative Linguistics, Swets & Zeitlinger, Vol. 9, No. 3, 215–231.
- SPEKTRUM, 1983. (Articles: Jak se vyhnout nebezpečí entropie. Nový konzervatismus. Současný problémy: Změna a kontinuita; pluralismus a fundamentalismus ve Spojených státech. Ženské hnutí, rodina a společnost.) Embassy of the United States of America, 43, Praha, 16–33.

WESTWOOD, J., 1988. The Pictorial History of Railroads. Gallery Books, Smith Publishers Inc., New York, 8–143.

WESTWOOD, J., 1994. Železnice - obrazové dějiny. Columbus, Praha, 8–143.