Totalitarian Dictionary of Czech

František Čermák
Institute of the Czech National Corpus, Charles University
frantisek.cermak@ff.cuni.cz

Abstract

Due to a decades-long experience with the Communist totalitatian regime, an attempt was made to capture main features of its language, and thus the substance of the time, by a corpus-based dictionary. The corpus, based on large samples of Communist newspapers and booklets, which is now accessible on the web, served as basis for a frequency- based dictionary of the time; it has been provided with statistics according to time periods, with pertinent collocations, appearance in various genres and, most of all, with comparative figures with today´s standard as it is to be found in a large contemporary corpus of Czech. This, by far the first systematic description of the period in any language, is complemented by a detailed study of the lexis of the time, many typical excerps, etc.

Keywords: corpus; frequency dictionary; dictionary of a period; Communist totalitarian vocabulary

1   Goal and Historical Background

Twenty years after the decline of the Communist regime in Czechoslovakia, the first attempt (Čermák et al. 2010) has been made, on the basis of corpus-data, to map the language of the period of over four decades of the Communist rule (1948-1989) in the Czech-speaking part of the country, known now as Czechia. Whether too soon or too late, the decision was precipitated by realization how fast people tend to forget the totalitarian time, young generation not knowing it at all. However, it seems that it is through the vocabulary of a period that one gets to know it best. Needless to say that the objective of solving the situation is best served by corpus-based data offering an eminent source of information about ways of manipulation of the whole society through its rather effective propaganda. There seem to be very few similar dictionaries delving deep into the language of a period objectively and systematically, based on corpus data. Thus the book is both a lexicographic and linguistic record of that time as well as a valuable historical and sociological description of what had long been a taboo, namely propaganda, which, in many respects, drew amply on Hitler and Goebbels from an earlier period of fascism.

2   Data and Their Preparation

Since it was not possible to cover linguistically all texts from the period, the *Totalitarian corpus* (available on the web at korpus.cz) had to be limited to over 15 million positions made up of 422 762 word forms which equals to enormous 164 815 lemmas. The data have been drawn painstainkingly from the major and all-embracing Communist daily *Rudé právo* (Red Right) from three historically critical but different periods of the second half of the year 1952 (6. 6. - 31. 12. 1952), i.e. from the peak of the most primitive communist propaganda, the second quarter of 1969, i.e. from after the Soviet and Russian occupation of the country in 1968 (1. 4. - 31. 4. 1969) and from the first quarter of 1977 (3. 1. - 31. 3. 1977), i. e. time of deep social recession and depression. The newspaper data amounting to some 10 million words have been complemented by a careful and typical selection of 91 propaganda books and booklets (about 5 million words) from the same period. All of this data have been manually scanned and corrected having to come to grips with several different spelling reforms and various standardisation processes. No attempt has been made to cover a later period when it was felt that the political regime was characterized by stagnation and inertia as well as gradual crumbling of its strongholds which culminated in 1989, i.e. its definite downfall. The data have been tagged and lemmatized which made it necessary to adapt the existing taggers and lemmatizers. Of course, the fact that the Totalitarian corpus is freely accessible on the Web now means that any further analysis is possible, although the primary goal has been to cover the period´s typical vocabulary. It may be reasonably hoped that the core of the vocabulary is captured in dictionary form being accompanied by an extensive analysis, complements and comparison with today´s language.

3   The Dictionary of the Communist Totality

The *Totalitarian Corpus*, based on a prior decision to choose data from three historical periods of eminent linguistic interest (and the availabale propagandist books and booklets), had to be further limited before dictionaries based on it have been compiled, both capturing only the top selection of the most important lemmas.

One of the main goals of the project was to point out to users what used to be typical in the vocabulary of the period (**A**). The tricky business of corpus comparison of the period with today´s standard language, which is several decades remote from the time recorded in a large contemporary corpus of Czech (SYN 2005, see korpus.cz), has been projected into frequencies (**B**) in one of two dictionaries compiled on this basis, namely the **Specific Totalitarian Dictionary** (*Dictionary* in the following). Because this dictionary, based on such comparison, has been achieved due to use of statistical methods it is highly selective in its nature capturing only those lexemes whose frequency has had a high degree of deviation from today's use. The dictionary is alphabetical giving a summary of absolute (ABS) and relative (REL) frequency (in square brackets) and partial frequencies of each of the three

periods examined (1952, 1969, 1977, given in ppm, i.e. parts per million). Numerical differencies between these three figures suggest a growing or falling use of the frequency of the words in the three periods. Obviously, many lexemes that have not seemed to be statistically different enough have been dropped and are not to be found here.

Where necessary, an additional brief explanatory note has been added identifying lexeme´s meaning. Typical collocations, that offer a more direct information about the use of the word are offered following a full dot • Should the lemma be out of use in today´s newspapers completely (i.e. not found in the contemporary 100-million corpus) it is preceded by a cross sign †, such as †**agrobiologický** (agro-biological). Positive, period-linked sign of a larger asterisk ★ standing before a collocation suggests a prominent use and specifically high frequency in the Communnist era, while the sign ✌ indicates, in contrast, a typical collocation of today, such as in **bezmezně** (boundlessly, infinitely, immensely) in „★ **b. oddáni** (*SSSR*) × ✌ *b. věřit* “ (*immensely devoted* to USSR at the time, as against to *infinitely believe* in today´s usage). The overwhelming devotedness to the Soviet Union, often found elsewhere, is captured in this example prominently, showing also how the usage of the verb *believe* (věřit) has eventually returned to normalcy now.

Collocations have been obtained by getting a combination of predetermined values of three association measures of lexemes having a frequency higher than 5, namely of MI-score, log-likelihood and Dice. It is specifically collocations that reveal the usage of heavily loaded words of the time such as fight/struggle:

boj (struggle, fight) • *boj dělníků* (workers´ struggle); *boj lidu* (za mír) (people´s fight for peace); *boj národů* (za mír) (fight of nations for peace); boj proletariátu (proletariat´s struggle); *ideologický boj* (ideological struggle); *ideový boj* (struggle of ideas); *organisovat boj* (pracujících) (organise fight of the workers); *podporovat boj* (národů) (support fight of nations); *revoluční boj* (revolutionary fight); *rozvíjet boj* (proti imperialismu) (develop struggle against imperialism); *rozvinout boj* (za mír) (develop fight for peace); *stávkový boj* (fight of those on strike); *třídní boj* (class struggle); *řídit boj dělnické třídy* (direct the fight of the workers´s class), etc.

On a closer inspection, it is evident that the society lived then in a state of frenzy feeling both endangered from outside and feeling that it must fight for almost everything, including even the most common, everyday things.

4   A Sample of the Specific Dictionary

Let us have a look at the Dictionary and some of its lemmas:

Adenauer   ABS 348, REL 27 ppm
[**83 ppm** | 1 ppm | 0 ppm] *Konrad Adenauer, kancléř NSR (1949-1963)*

Adolf   ABS 92, REL 7 ppm

[**6 ppm** | 12 ppm | 5 ppm] • Adolf Hitler

agent   ABS 1292, REL 100 ppm

[**263 ppm** | 22 ppm | 24 ppm] • agent buržoazie; agenti imperialismu; imperialistický agent;
†titovský agent

agitace   ABS 894, REL 69 ppm

[**104 ppm** | 12 ppm | **85 ppm**] • †agitace komunistů; názorná agitace; osobní agitace;
politická agitace; účinnost ekonomické propagandy a agitace

agitátor   ABS 811, REL 63 ppm

[**166 ppm** | 2 ppm | 23 ppm] • †kolektiv agitátorů; †příprava agitátorů; †seminář agitátorů;
†schůze agitátorů; stranický agitátor

agrese   ABS 1213, REL 94 ppm

[**188 ppm** | 50 ppm | 50 ppm] • hitlerovská agrese; imperialistická agrese;
(definice) pojmu agrese; politika agrese

americký   ABS 15089, REL 1170 ppm

[**2527 ppm** | 601 ppm | 476 ppm] • americký agresor; američtí barbaři; americký businessman;
americký imperialismus; americký katan; americký monopolista; americké okupační úřady;
americký okupant; †američtí podněcovatelé války; †americká soldateska; američtí supermani;
provokace amerických imperialistů

angažovaný   ABS 275, REL 21 ppm

[0 ppm | **38 ppm** | **26 ppm**] • angažované země; †schůzka neangažovaných zemí

armáda   ABS 6993, REL 542 ppm

[**1067 ppm** | 340 ppm | 260 ppm] • lidová armáda; (spartakiáda) spřátelených armád;
západoevropská armáda; Svaz pro spolupráci s armádou (*viz Svazarm*)

balistický   ABS 16, REL 1 ppm

[**0 ppm** | 1 ppm | 2 ppm] ★ (mezikontinentální) b. raketa × ✌ b. expertíza

bdělý   ABS 86, REL 7 ppm

[**18 ppm** | 1 ppm | 2 ppm] ★ politicky b. × ✌ b. stav

bezmezně   ABS 23, REL 2 ppm

[**4 ppm** | 0 ppm | 1 ppm] ★ b. oddáni (SSSR) × ✌ b. věřit

5   Basic Dictionary of the Communist Totality

To balance this key, though highly (i.e. statistically) prominent, dictionary, a second dictionary, that of Basic Dictionary (Základní slovník) has been compiled that is complementary to the Specific Totalitarian Dictionary. **The Basic Dictionary** is represented by almost ten thousand most frequent lemmas from the Totalitarian Corpus (9994, i.e. those having more than 63 occurrences). Thus a comparison of the specific period-related vocabulary obtained statistically and that of the systematic full dictionary is enabled, the latter enabling to find there also those words that were used but were not statistically interesting enough. This dictionary, based on descending absolute frequency ordering, shows all lemmas provided within the given frequency limits. The sum-total of frequency for each lemma is broken into three different frequency figures for each of the three periods covered. For cross-reference purposes, those lemmas given also in the Specific Totalitarian Dictionary are marked by asterisk.

Lemma	Overall	1952	1969	1977
a	498450	166271	130175	202004
v	422528	129014	126641	166873
být*	339917	110927	101912	127078
se	225925	67849	72378	85698
na	214183	70624	66385	77174
ten	120081	36929	36986	46166
s	106963	31667	33946	41350
který	104580	36545	30776	37259
že	99118	34209	31456	33453
z	95681	27971	31637	36073
strana*	52942	22081	12821	18040
všechen	51970	22246	11180	18544
rok*	49349	13494	15128	20727
práce*	44491	17044	8300	19147
jeho	39773	11046	11304	17423
aby	37883	17291	8835	11757
sovětský*	36644	22690	4495	9459
socialistický*	35990	7450	5373	23167
nový*	34985	13992	8061	12932

Looking at the first 30 most frequent lemmas of the Basic Dictionary it is just appaling to see what kind of priority some of its words used to have in the Communist-ruled society at the time (although some words have been left out). Thus, the very first noun is *strana* (party, frequency 22) followed by *práce* (work) while the very first adjective used then was *socialistický* (socialist, fr. 29), followed by *nový* (new). The prominent presence of nouns does not need much comment telling what had been dominant at the time: Communist party and work. Likewise, the adjective socialist is to be expected here while the particular adjective new would deserve a special study, not possible here, since every second thing, notion, aspect of life coming from the past had to, on ideological grounds, be rejected and replaced by something new, i.e. *Communist*, most often *Soviet-inspired*, since Soviet Union had been presented as a new Promised Land serving as a model and exemplary goal to be followed by everyone and serving as an official inspiration for everyone. Some extensive collocations of this type showing this in detail are added in the Supplement to the Dictionary.

6   Additional Features of the Dictionary

Next to the two dictionaries, there is a large analytical study attached (*Slovník komunistické totality: lexémy, nominace a jejich užití, Dictionary of the Communist Totality: lexemes, nominations and their Use*) preceding both dictionaries, based on the data added, scrutinizing and researching typical words and collocations more deeply paying special attention to principles of the Communist propaganda. Some attention is paid to comparison of joint features of the fascist propaganda and the Communist one and a semiotic analysis of how certain lexemes were used for manipulation is shown.

Since all data have been published only after they underwent a severe censorship, an attempt is also made to counterbalance this *officious (official)* language by a sample of the most important words of the *unofficial* language used by normal people and heard on the street that were not subject to the official parlance of the time and never printed. It is shown that beginnings of this second, unofficial language or rather vocabulary was born and used profusely wherever people were in private and not followed by the secret police; thus a parallel vocabulary of synonyms was born (e.g. *strana-partaj, komunista-komouš,* etc). The study is complemented by a number of typical samples of texts, vocabulary and collocations, showing, for example, the blind reliance on anything Soviet whose meaning has changed from a proper name so much as to just practically designate only what was „good, best, exemplary".

7   Notes and Conclusions

Slovník komunistické totality (Dictionary of the Commununist Totality) maps the gap between a general dictionary and one of an epoch or period enabling, in this case, comparison of two comparati-

vely close time periods, an attempt rather rare in lexicography, not to speak of its wide sociological and political impact.

The fact that there is a substantial specific corpus behind it (which is not the case of German, Polish or Russian where similar books have been published) is both a possibility and offer to to anyone who wishes to study the period language more widely, and, should other dictionaries arise in other languages (from the period of Communism and Fascism primarily) it might become a contribution to a broader understanding of political and social movements spanning substantial parts of Europe not long ago - and not only that.

8   References

Slovník komunistické totality (Dictionary of the Commununist Totality)**,** eds. František Čermák, Václav Cvrček, Věra Schmiedtová, further coauthors Jan Kocek, Dominika Kováříková, Karel Kučera, Renata Novotná, NLN Praha 2010

Arendtová, H. (2000). *Původ totalitarismu I a II.* Oikoymenh, Praha.
Bartošek, K., J. L. Margolin (1999). *Černá kniha komunismu: Zločiny, teror, represe;* Paseka, Praha, Litomyšl.
Bralczyk, J. (2001). *O języku polskiej propagandy politycznej lat siedemdziesiątych.* Wydawnictwo Trio, Warszawa.
Brenner, Ch. (2008). *Znormalizovaný totalitarismus? Paradigmata výzkumu socialismu I,* Ročenka textů zahraničních profesorů; FFUK, Praha.
Brousek, A. (1987). *Podivuhodní kouzelníci, čítanka českého stalinismu v řeči vázané z let 1945-1955.* Rozmluvy, Londýn.
Cvek, B. (2005). *Proč a v čem je komunismus vlastně totéž co nacismus*; Britské listy 26. 2. 2005. Dostupné z WWW: <http://www.blisty.cz/art/22182.html >.
Džilas, M. (1977). *Nová třída, kritika soudobého komunismu.* Demos, Curych.
Fidelius, P. (1983). *Jazyk a moc.* Edice Arkýř, Mnichov.
Fidelius, P. (2000). O totalitním myšlení, In *Kritické eseje.* Torst, Praha.
Głowiński, M. (1991). *Nowomowa po polsku.* Wydawnictwo PEN, Warszawa.
Hochel, B. (1991). Totalita v jazyku (a v nás). In *Kultúrny život,* 25, Bratislava.
Jánský, P. (2004). *Totalita slovem, písní a obrazem (Dějiny hrůzovlády KSČ).* Nakladatelství Music, Cheb.
Klemperer, V. (2002). *Deníky 1933-1941, Chci vydat svědectví.* Paseka, Praha, Litomyšl (*Victor Klemperer: Die Tagebücher 1933–1945. Kritische Gesamtausgabe. CD-ROM. Berlin 2007*).
Klemperer, V. (2003). *Jazyk Třetí Říše – LTI: poznámky filologovy.* Nakladatelství H&H, Jinočany.
Kartous, B. (2004). *Může totalita zahltit celou civilizaci?* Britské listy 10. 11. 2004. Dostupné z WWW: <http://www.blisty.cz/art/20532.html>
Macura, V. (1992). *Šťastný věk.* Pražská imaginace, Praha.
Nowak, P. (2002). *Swoi i obcy w językowym obrazie świata: język publicystyki polskiej z pierwszej połowy lat pięćdziesiątych.* Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin.
Mokijenko, V. M., T. G. Nikitina (1998). *Tolkovyj slovar jazyka sovdepii.* Sankt-Peterburg.
Mokijeno, V. M. (2000). *Novaja russkaja frazeologija.* Opole.
Popper, K. (1994). *Otevřená společnost a její nepřátelé I. a II.* Oikoymenh, Praha.
Pisarek, W. (1972). *Frekwencja wyrazów w prasie.* Ośrodek Badań Prasoznawczych, Kraków.
Pisarek, W. (1976). *Język służy propagandzie.* Ośrodek Badań prasoznawczych, Kraków.
Pisarek ,W. (1983). *Analiza zawartości prasy.* Ośrodek Badań prasoznawczych,Kraków.

Röhrich, A. (2008). *Ideologie, jazyky, texty: Analýza a interpretace textů Rudého práva z roku 1953 a 1975 a Práva z roku 1997.* Bor, Liberec.

Šebesta, K. (2001). Studovat jazyk totality. In: *Institucionalizace (ne)odpovědnosti: Globální svět, evropská integrace a české zájmy.* Karolinum, Praha.

Šlosar, D. (1995). Jazyk totality a jazyk dneška, In *Spisovná čeština a jazyková kultura 1993.* FF UK, Praha.

Totalitarismus 3 (2007). Eds. I. Budil a T. Zíková. FF ZU, Plzeň.

Totalitarismus 4 (2008). Eds. I. Budil a T. Zíková. FF ZU, Plzeň.

Vítězové? Poražení? I. díl, Disent v období tzv. normalizace (2005). Eds. M. Vaněk a P. Urbášek, Prostor, Praha.

Vítězové? Poražení? II. díl, Politické elity v období tzv. normalizace (2005). Eds. M. Vaněk a P. Urbášek, Prostor, Praha.

Wierzbicki, P. (1986). *Struktura klamstva*, Glos, Warsawa.

Włodek, M. (2002). *Sondy do jazyka totality.* diplomová práce, FFUK, Praha.