

Hodný, zlý a ošklivý (*The Good, the Bad and the Ugly*)

The representation of three minority groups in printed media
discourse from the Czech Republic

Irene Elmerot

Institutionen för slaviska och baltiska språk, finska, nederländska
och tyska

Kandidatuppsats 15 hp /Bachelor thesis 15 HE credits

Tjeckiska/Czech

Tjeckiska IV (91–120 hp) /Czech IV (91–120 credits)

Läsåret/academic year 2016–2017

Handledare/Supervisor: Tora Hedin

Svensk titel: Den gode, den onde, den fule – tre minoritetsgruppers
representation i tjeckisk mediediskurs



Stockholms
universitet

Den gode, den onde, den fule

– tre minoritetsgruppers representation i tjeckisk mediediskurs

Irene Elmerot

Sammanfattning

Syftet med detta arbete är att göra en kvantitativ, korpusbaserad undersökning av den språkliga andrafieringen av tre minoritetsgrupper: romer, vietnameser och ukrainare, i den tjeckiska mediadiskursen under 15 år, samt att få ett omfattande, representativt resultat genom att jämföra neutrala, positiva och negativa adjektiv som står intill sökorden. Till teoretisk grund ligger hur språket hjälper till att bygga och förstärka maktstrukturer samt hur korpusökningar efter kollokationer och närliggande ord kan tydliggöra en språklig andrafiering. Här används en kvantitativ metod för att besvara analytiska frågor om dessa benämningar. Materialet som ligger till grund för analysen är SYN version 4 i det tjeckiska Nationalkorpuset – i sin helhet består det av 275 miljoner meningar. Det verkar inte tidigare ha utförts någon sådan undersökning på ett så stort material, även om några forskare har använt liknande metoder.

Resultatet bekräftar att de olika grupperna beskrivs på olika sätt, och är, genom det stora källmaterialet, ett bevis på hur språket i mediadiskursen speglar diskursen i samhället i stort.

Nyckelord

korpusundersökning, språklig andrafiering, mediadiskurs, tjeckiska kollokationer, stereotyper, frekvensundersökning

Keywords

corpus analysis, linguistic othering, media discourse, collocations in Czech, stereotypes, frequency analysis

Table of contents

1. Introduction	1
2. Aim and focus of the study	2
3. Theoretical framework	3
4. Previous research	5
4.1. Linguistic research.....	5
4.1.1. Linguistic othering – collocations and corpora	5
4.1.2. Critical discourse analysis	6
4.2. Non-linguistic research.....	7
4.2.1. Research on Roma	8
4.2.2. Research on Vietnamese.....	8
4.2.3. Research on Ukrainians	9
5. Socio-political background of Roma, Vietnamese and Ukrainians in the Czech Republic	9
5.1. Roma.....	10
5.2. Vietnamese	11
5.3. Ukrainians	12
6. Questions	12
7. Source material and method.....	13
7.1. Work steps	17
8. Results	19
8.1. The most frequent adjectives	19
8.1.1. Adjectives before Roma	21
8.1.2. Adjectives before Vietnamese.....	21
8.1.3. Adjectives before Ukrainian.....	22
8.2. The classification of the adjectives	22
9. Concluding discussion	24
References	27
Internet sources.....	29

1. Introduction

This study concerns the othering of three minority groups in the Czech Republic. The idea is to make this a part of a larger research of linguistic othering, based on sources from different times in the Czech Republic, and it is a follow-up to my previous study comparing two different terms, *Cikán* and *Rom*, for the same minority group (Elmerot 2016).

The source material for this study is the corpus SYN, version 4, with texts from 1989 till 2015, described in more detail in the chapter on [source material](#). Since I look for adjectives adjacent to the keywords for these minority groups, this is a study of collocations¹ – words “repeatedly found in the close vicinity of a node word in texts” (Teubert & Čermáková 2007, 139) or “the above-chance frequent co-occurrence of two words within a pre-determined span” (Baker et al. 2008, 278). The term is here used for the words statistically proven to be the most frequent.

The title of this thesis comes from the 1966 Italian film “Il buono, il brutto, il cattivo” by Sergio Leone, where three men engage in the same thing as many of the people described in the source material for this study – shooting and committing awful deeds.

I would like to thank my tutor, head of the department Tora Hedin, for guiding me *inter alia* through the garden paths of discourse analysis, as well as senior lecturer Klas Rönnbäck for helping me with the statistics and structure.

¹ *collocation* the act or result of placing or arranging together; specifically: a noticeable arrangement or conjoining of linguistic elements (as words). “To save time” and “make the bed” are common collocations. *Collocation*. (n.d.). <https://www.merriam-webster.com/dictionary/collocation>

2. Aim and focus of the study

The aim of this study is to quantitatively examine the linguistic othering of Roma, Vietnamese and Ukrainians in Czech media discourse during the years after the Velvet Revolution in 1989, and also to get a comprehensive and representative result by comparing neutral, positive and negative adjectives adjacent to the keywords *Rom*, *Vietnamec* and *Ukrajinec*. To fulfil this purpose, a large enough source material must be used, not only a few articles or newspapers. Then a quantitative study has to be made, followed by an analysis of the adjectival frequency results and a comparison with previous research to show in what way this result fills previous gaps.

The largest minorities in the Czech Republic are, as far as the Czech Statistical Office knows, Slovaks, Ukrainians, Vietnamese, Russians² and Roma³. This study focuses on three of these groups. Many non-linguists have – as shown below under the chapters on [background](#) and previous [research](#) – written about how the Roma, Vietnamese and Ukrainians are treated in the Czech media, and they also give an interesting background to this study, especially for the classification of stereotypical adjectives.

The general focus of this study is to look at how the construction of identity for the people at issue is made in the mainstream written media. Othering is a subject that defines how power is (mis)used, and this study focuses specifically on how language use makes it obvious.

² https://www.czso.cz/documents/11292/27914491/1512_c01t01.pdf/7e9ea54d-13c8-4a41-bffa-bfe518d66ddc?version=1.0

³ The actual number of Roma persons in the Czech Republic is not clear.
<http://www.romea.cz/en/news/czech/patrik-banga-i-m-for-collecting-ethnic-data-in-all-areas-but-the-majority-can-t-tell-who-is-romani-2>

3. Theoretical framework

The theoretical starting point for this work is that power relations in a society are reflected in that society's general media, and that these media's language use contributes strongly to the receivers' worldview, in some cases even helps to construct it. Hence, the language of popular (as in "directed to a general audience and therefore very often and widely read or listened to") media is a power that forms general stereotypes that in turn are reflected in the discourses of society as a whole.

Othering is a term that may need an explanation, since Encyclopaedia Britannica, Merriam-Websters nor the American Heritage Dictionary include the word. There is, however, an explanation in *The New Fontana Dictionary of Modern Thought* on page 620:

The term Othering describes the reductive action of labelling a person as someone who belongs to a subordinate social category defined as the Other. The practice of Othering is the exclusion of persons who do not fit the norm of the social group, which is a version of the Self.

Minority groups in a society are such "social categories", and they are the subject of othering, whether or not this term is used (cf. Fairclough 2015, 40–41; Loomba 2015, 115).

Othering is also a new term for an old concept (Loomba 2015, 112–113). It has been discussed in many research fields, from how Christians othered Jews (the picture of a greedy person with a large nose is perhaps way too well-known) in different times and places throughout history, to how modern-day politicians try to blame refugees for different crimes just because they are from an "other culture".

To show what the discourse then is in a society, to pinpoint how it's built with a certain language use that is often repeated, is therefore as intriguing a study as ever. The most famous example from previous research theorists in the Western world is perhaps Edward Said's book *Orientalism* (first written in the mid-1970s, according to Said's own "Acknowledgments"), but as long as there has been research on minorities, there have been theories about the other. But, as Edward Said put it in his 1995 edition's Afterword (Said 1995, 332), "Each age and society re-creates its 'Others'. /.../ In short, the

construction of identity is bound up with the disposition of power and powerlessness in each society, and is therefore anything but mere academic wool-gathering.”

When Norman Fairclough first wrote *Language and Power* in 1989, one reason was to fill the gap that he thought sociolinguistics had missed, in “the rich and complex interrelationships of power and language” (Fairclough 2015, 1). When minority groups like the ones in focus for this study are mentioned in the media discourse, it is a use of what Fairclough calls “hidden power” (Fairclough 2015, 41), an expression that may be even more true today than in 1989, with anonymous haters affecting the media discourse. The frequent but widely dispersed stereotypical and negative expressions and collocations are examples of this “hidden” power, that may not be visible at once, but slowly enters the general discourse in a society.

Masako Fidler and Václav Cvrček mention how we form our conception of the world through words that are often collocated, i.e. found together, like the phrase *illegal immigrant*. I’d like to make use of their theoretical framework: Language use that we consistently are exposed to creates a pattern that in turn forms our conceptions and expectations; if we then often read two words together, we eventually make it into a collocation, something that is closely tied together (Fidler & Cvrček 2015, 198–199).

Making use of a large corpus to see this will create a better result than only reading a few articles.

Using corpora helps me to use these theories, as it would have been hard to do so before there were corpora of this size and computers like the ones we have today (McEnery & Hardie 2012, 1).

4. Previous research

4.1. Linguistic research

4.1.1. Linguistic othering – collocations and corpora

Othering has been the focus of many studies through the years – Edward Said’s *Orientalism*, Ania Loomba’s *Colonialism/Postcolonialism* and Norman Fairclough’s *Language and Power* are cited above, in the chapter on [theoretical framework](#). One of the few researchers who have done something similar to this study, researching fixed expressions or collocations, is Masako Fidler. In her 2016 article on othering in the Czech Republic, she writes about her intention to “find a more automatic mental representation” of the “others” that she studies. Her take on collocations or adjectives that often are put together with certain keywords, is that “Dispersed and repeated, such expressions are likely to cumulatively reinforce a certain mental representation; at the same time they are so automatic they may not be noticeable to the reader.” (Fidler 2016, 38–39). This mental representation is an aspect of othering.

Even the Roma people’s language, Romani, is “othered” in that it is not at all mentioned as a collocation to the word *mluvící* (“speaking”), according to Fidler’s quantitative study of collocations and a keyword analysis on a source material from the Czech National Corpus. In that article, she finds that Romani, which is in fact one of the four officially protected languages of the national minorities in the Czech Republic, is not even one of the top 30 collocates in her source material from 1989 to 2009 (Fidler 2016, 56).

A Czech collocation that is attached mostly to Roma is “unadaptable”, *nepřizpůsobivý*. In my previous study (Elmerot 2016), I also concluded that it is more often used with the keyword *Cikán* than *Rom*, at least adjacent to the keyword in question. It may originally have been a euphemism for harsher, more negatively charged words, but is in the Czech Republic of today used to mean Roma (Slavičková 2015, 74) or other non-Czech, law-breaking persons. The negated form (“un-”) is in a corpus search

much more frequent than its positive counterpart *přizpůsobivý*. It is also used in different kinds of text, not only journalism, but it is by far more common in news papers and magazines, and when it's found in other categories, it's still very often collocated to Roma (Slavičková 2015, 75). This is an example of an othering collocation that sums up the idea of a certain other without being too explicit.

4.1.2. Critical discourse analysis

Critical discourse analysis, sometimes abbreviated CDA, has in the past couple of decades expanded from small, qualitative studies to be used also in larger projects. An article from one such project was published in 2014, where the researchers analysed anti-minority rhetoric with “the aim of extending existing linguistic studies (that have also focused mainly on Roma) in the Czech Republic”

(Slavičková & Zvagulis 2014, 155). Their conclusion is that the Roma are a special target for this anti-minority rhetoric (idem, 153) but that many officials pay “lip service” to formally respect different groups, so that it looks like there is no value-laden language use around. Unfortunately, the researchers do not mention the term *Cikán* in the entire article, which might have been a comparison worth making – and Slavičková later does in her 2015 article. The idea that media and society today too often mention minority people primarily when it comes to social problems – from lack of education to different kinds of abuse – is reflected in their article (Slavičková & Zvagulis 2014, 153), and this is something that would, if true, also show through the negative adjectives in this study. In addition, they have concluded that especially Roma and Vietnamese rarely appear as newsmakers except in negative contexts (Slavičková & Zvagulis 2014, 160).

There is also a discourse analysis in *Obraz Romů v středoevropských masmédiích po roce 1989* (*Presentations of Roma/Gypsies in the Central European Media after 1989*) edited by Kamila Karhanová, Jiří Homoláč and Jiří Nekvapil, where they, too, conclude that the Roma are presented in a negative context, and try to pin-point which “rhetorical devices and genres” are normally used in the spoken and written media (Karhanová, Homoláč and Nekvapil 2003, 3). Their work does, however, not mention the focus of my study: which exact wordings and terms that were used in the media, although the writers Anna Šabatová, Jiří Homoláč and Kamila Karhanová state in their chapter (page

100; page 102 for the English version) that where a Romany is the crime victim, journalists take it as an excuse to mention negative stereotypes, even in comparably neutral texts.

In my pilot study (Elmerot 2016) I could confirm that there are more often prejudiced collocations to the term *Cikán* than the term *Rom*, which is why I'd now like to compare the latter term with the generally more appreciated Vietnamese and Ukrainians.

I found one article that deals with Ukrainians and Czechs from a linguistic perspective, comparing oral communication patterns between the two peoples, and in this, authors Podolyan & Nývltová say that the Ukrainian migrants that have moved to the Czech Republic in recent years have been able to assimilate “massively quick” (Podolyan & Nývltová 2012, 313). They argue that the reason is the two nations' joint linguistic Slavic and (for Western Ukraine) political Austro-Hungarian roots.

Previous research from different, non-linguistic, fields (presented below) also helps me get a better picture of what words are really stereotypical, which creates a more stable ground for my classification.

4.2. Non-linguistic research

There seems to be much discomfort among Czech citizens with immigrants. Above-mentioned anthropologist Zalabáková mentions several articles that I've found useful to read as a background. Among the academic research, she also quotes a poll performed in 2001, where 81 per cent of the respondents agreed or mostly agreed with the statement “The increased number of foreigners contributes to the increase of criminality and terrorism.”⁴ as well as a poll from 2009, where 74 per cent agreed wholly or mostly with the statement “Foreigners that stay for a long time in the Czech Republic contribute to the criminal growth.”⁵. The latter statement actually includes all three groups

⁴ *Zvyšování počtu cizinců přispívá k růstu kriminality a terorismu.*

⁵ *Cizinci pobývající dlouhodobě v České republice se podílejí na nárůstu kriminality.*

that are in focus for this study (Zalabáková 2012, 84–85), although they arrived in very different times.

4.2.1. Research on Roma

The Roma have, as can be deduced from [the background chapter](#), a peculiar position in the Czech Republic – they're often mentioned in the context of “distant others”, people from other countries and continents (Zalabáková 2012, 83), despite having lived in the country for centuries.

In the 1900s before 1989, the Roma were othered in the public discourse, as shown by e.g. Alamgir (2013, 74–75), although the official idea from 1948 onwards was that they were indeed able to change, not “unadaptable” (*nepřizpůsobivý*) (cf. [above](#)). They were then also called “backwards” (*zaostalý*), and were, as Alamgir shows (2013, 81), often compared with the Vietnamese workers, and then often with more pejorative words (Alamgir 2013, 82).

Not until the 1990s do we find oral and written sources where Roma are mentioned in a serious and respectful way, and where journalists and other writers use the term Roma instead of Gypsy (Schneeweis 2009, 147).

4.2.2. Research on Vietnamese

Hantonová (2013, 67–71) has researched what themes were most frequent in four Czech newspapers in 2012, and even if it's a small source material, it may give a hint about the general picture. In these papers, the favourite subjects were the Vietnam war(s), economics/business and politics, although the daily paper *Právo* also adds 7 articles each, out of 60 in total, on the Vietnamese minority and on criminal reports (“krimi”).

Alamgir (2013, 79) notes that in the press from the 1980s, the Vietnamese were described in an infantilizing way, which echoes in some of the adjectives found in this study (*drobný*, *maličký*, *útlý*, all meaning “tiny” or “petite”, as well as *šišlající*, meaning to talk “baby talk” or lisp).

4.2.3. Research on Ukrainians

Many researchers (e.g. Zalabáková 2012, 81–82; Leontiyeva 2006, 91) have found that Ukrainians are treated differently than other minorities in the Czech republic, simply because they come from a neighbouring country, and a Slavic-speaking, at that. When they appear in the papers, however, it's very often as drunkards, murderers or thieves/robbers (Zalabáková 2012, 84–85; Součková 2015, 68; Leontiyeva 2004, 27).

Something worth comparing to the results of this study is that of these three groups, the Ukrainians are also mentioned in a religious context in previous research on minorities in the Czech Republic (Leontiyeva 2004, 26; idem 2006, 37; Součková 2015, 69, 72).

After the events in 1989–1990, there seems to be two groups of Ukrainians depicted in the Czech media: “guest working” (i.e. drinkers and possible criminals) and “non-problematic” (Leontiyeva 2004, 27), and we'll see if this duality is obvious from the results of this study, too.

5. Socio-political background of Roma, Vietnamese and Ukrainians in the Czech Republic

The groups in this study have been living in comparatively large numbers in the Czech Republic for decades (Vietnamese, Ukrainians) and centuries (Roma). According to the Czech Statistical Office,

both Vietnamese and Ukrainians came in larger numbers after 1989, and constitute today about 0.3 and 0.5 per cent respectively of the population⁶.

5.1. Roma

The Roma have been in the Czech lands since the year 1416 (*Ottův slovník naučný* 2892, 366), or possibly even before that. During the National Revival in the 19th century, they emerged into popular books and songs as an exotic and culturally active people, of which the most famous Czech example is Karel Hynek Mácha's *Cikáni*. The descriptions of today are, as we can see in this study, much less romantic, proud and colourful.

The amount of Roma in the Czech Republic is difficult to assess, since many of the persons with Roma family or ancestry try their best to avoid being registered as such (Doubek; Levínská; Bittnerová 2015, 132–133). It's therefore also difficult to give a certain statistical background as to how many live in the Czech Republic and therefore about their life. The web portal Romea.cz has tried to put some facts together, aided by the Museum for Roma culture and history⁷, and they state that in previous centuries, Roma have been considered a legal prey for anyone – it wasn't illegal to kill a Roma person. In 1925 there was a Czechoslovak census counting 62,192 Roma in Slovakia, 1994 in Moravia, 579 in the Czech lands and 79 in Silesia. There was in 1927, as a result, a law prohibiting Roma from moving around freely. A decade later, the adults got a "Gypsy identification paper" that they had to carry at all times. At the same time, some Roma, however, also started attending higher education and schooling was seen as a way of integrating them into the Czechoslovak society⁸. After 1945, many Slovak Roma moved to the Czech and Moravian parts of the Republic, due to the previous genocide of these groups. During the years 1948–1989, the authorities officially tried more ways of improving the life of the Roma by schooling and health information, but also demanded that they ignore their Roma identity in favour for their Czechoslovak identity. After 1989, there was a political

⁶ <https://www.czso.cz/documents/10180/20551765/170223-14.pdf>

⁷ <http://skola.romea.cz/cs/historie/>

⁸ <http://skola.romea.cz/cs/historie/romove-v-prvni-ceskoslovenske-republice/>

Roma party that got into the parliament for a few years. Still, not all Roma have fulfilled an education even today, and many are unemployed and poor, even if the exact numbers are hard to get⁹.

5.2. Vietnamese

The Vietnamese began immigration in larger numbers to the Socialist Republic of Czechoslovakia, ČSSR, in the 1950s after some diplomatic relations had started between the two countries (Hantonová 2013, 8), and an agreement was written on scientific-technical co-operation (Hantonová 2013, 9; Alamgir 2013, 71). From the beginning they were perceived as quiet and hard-working (Hantonová 2013, 10). One reason for this may be that the ones that arrived during the 1980s did not get as much schooling in the Czech language as their predecessors (Alamgir 2013, 72; Hantonová 2013, 10) – although today, the Vietnamese are mentioned specifically as having a generally high educational level¹⁰.

When the Velvet revolution occurred, there were almost 30,000 Vietnamese in the ČSSR (Hantonová 2013, 11), whereas they are supposedly today some 48.000 permanent residents that still have a Vietnamese citizenship¹¹, and some 60.000 altogether¹². After 1989, the new Czechoslovak republic (Česká a Slovenská Federativní Republika, ČSFR) continued the transnational agreements, and after 1993, the Czech Republic and the Socialist Republic of Vietnam immediately got diplomatic relations (Hantonová 2013, 12).

Of the Vietnamese in the Czech Republic, about half of the estimated 60,000 Vietnamese residents in the Czech Republic have the right to a permanent residency. Still, not many of them have a Czech citizenship, which would, among other things, allow them to become more active in political elections

⁹ <http://skola.romea.cz/cs/historie/romove-v-ceske-republice-po-roce-1989/>

¹⁰ <https://www.czso.cz/documents/10180/20551765/170223-14.pdf>

¹¹ https://www.czso.cz/documents/11292/27914491/1612_c01t11.pdf

¹² https://www.czso.cz/csu/czso/cizinci_v_cr_vietnamci_u_nas_zakorenili20120214

and parliamentary politics. The Vietnamese have no parliamentary representation (unlike the Roma had for a while), despite their mainly 60 years of residency (Slavičková & Zvagulis 2014, 157–158).

5.3. Ukrainians

The Ukrainians are a Slavic people from a country close to the Czech Republic (and neighbour to Slovakia) that has been the focus of a few studies lately. Today, the Ukrainians are often considered to be guest workers, but have in recent years more often stayed longer than just a while, and even settled in the Czech Republic (Leontiyeva 2014, 63). In December 2016, there were some 110.000 Ukrainian citizens in the Czech republic, of which 81.000 had a permanent residence in the country¹³. However, the history of Ukrainians in the Lands of the Bohemian Crown is older. The most famous in a historic perspective may be the Cossacks that were part Ukrainians, but the biggest influx began in the 20th century with Ukrainian academic institutions being established in the Czech lands (Součková 2015, 67). 1919–1939, the south-westernmost part of present day Ukraine, the Carpathian Ruthenia, was also a part of Czechoslovakia.

6. Questions

The power relations and structures of a society are reflected and built in the general media discourse, and the language use in this discourse helps to maintain those relations and structures. The aim of this study is described [above](#), and to reach that aim, the following questions will be posed:

1. To what extent are these minority groups described in a stereotypical manner?
2. Are the “culturally” more distant Roma and Vietnamese more pejoratively othered than the Slavic Ukrainians?

¹³ https://www.czso.cz/documents/11292/27914491/1612_c01t11.pdf

7. Source material and method

In this study, the source material is the currently latest version of SYN, version 4¹⁴, from the Czech National Corpus, abbreviated ČNK. The searches were conducted mainly during February 2017. The SYN version 4 is a synchronic, non-representative (since it mainly reflects journalistic texts), referential corpus consisting in its entirety of about 14 million texts made up of 275 million sentences. The SYN series are monitor corpora (cf. McEnery & Hardie 2012, 6), and this means that it is well-made for searching large volumes of text with a statistical method to get an overview of the every-day written use of different collocations – “as the corpus grows, we might assume that any skew in the data naturally self-corrects, since there is no *consistent* skew in the data input” (McEnery & Hardie 2012, 7).

To be specific, the corpus consists of the national daily newspapers *Mladá fronta DNES*, *Lidové noviny*, *Právo*, *Hospodářské noviny*, *Blesk*, and *Sport*, as well as regional daily newspapers (chiefly *Deníky Bohemia* and *Moravia*) and non-specialized magazines (*Reflex*, *Respekt*, *Týden*) from the years 1990–2015.

This material was chosen because it seems to be the best for the aim of this study: to be able to make a quantitative study of what adjectives are the most frequent before the chosen minority groups' names, so as to get a representative, comprehensive picture of the language use in the most commonly read papers and magazines, and analyse to what extent these groups are treated stereotypically and “othered”. To use a smaller material or to do a study on only a few articles or papers would not fulfil this aim.

Collocations are chosen as a means to this end, because they tend to lead one's mind to their other half, like when we read “eclipse”, there are only a few words that come to mind, and when a Czech

¹⁴ <http://wiki.korpus.cz/doku.php/en:cnk:syn:verze4>

reader sees the word “socialně”, one of the first things that comes to mind is “slabý”, which then forms the collocation “socially deprived”. They are also a means for this method to statistically treat a large material.

To get relevant answers to my questions, the method used is a combination of statistical analysis of collocations – in this case which adjectives that were most often written adjacent to the specific people’s names from a large corpus and a analysis of the printed media discourse. A statistical analysis has to be made when the source material is as large and comprehensive as this. To make sure the results reflect the real world and an actual language use, a large corpus is a very suitable source material (cf. Slavičková 2015, 70). Another aspect is what Lüdeling & Kytö wrote about the combination of socio- and corpus linguistics, that corpus linguists must learn from socio-linguists to get a more complete research done (Lüdeling & Kytö 2009, 1122).

A statistical analysis of frequency is the best way to get something useful out of even the largest corpus. In this study, the statistical analysis of neutral, positive and negative adjectives adjacent to certain peoples gives us good data to analyse the linguistic othering in public, written discourse. It gives me the possibility to point out the contrast between positive, negative and neutral words in a large sample of the most popular and well-spread papers and magazines of the Czech Republic, and to pinpoint the power structures they reflect.

Masako Fidler stated that she wanted “to use language as a pointer towards a conceivable mental representation of the Others, underlying the image of national minorities and immigrants as perceived by the majority population in the Czech Republic” (Fidler 2016, 38). This is a basis also for this study, but with a slightly different method. Fidler & Cvrček (2015, 199) also wrote that “Statistical approaches help to reduce researcher bias and complement the qualitative analysis of texts”.

An analysis of the results of a corpus search is in this work chosen to prove how these expressions, dispersed but repeated, reflect a discourse that may strongly contribute to the worldview of the source

material's readers. By searching corpora (the SYN series, version 4, from the Czech National Corpus, www.korpus.cz) I here get an overall view of the language use of the Czech media on the chosen key groups. Whether this is to be classified as corpus linguistics or simply "corpus methods in linguistics", as McEnery & Hardie put it, is up to the reader to decide. (McEnery & Hardie 2012, xiv). The reader that wants to know more about SYN may use their wiki or read (Cvrček, Čermáková, Křen, 2016, 83–101) about SYN in general (idem pp. 83–90) and SYN15 in particular (idem, pp. 91–99). I may also need to add here that I've chosen this method for the sake of representativity, since I am of course aware that a corpus, however large, is only a fraction of the language total. Or, as McEnery & Hardie (2012, 15) put it, "We can only seek total accountability relative to the dataset that we are using, not to the entirety of language itself."

My keywords have been chosen on the basis of having an adjective on the place immediately to the left of the keyword, in order to get a meaningful, statistical result. One interesting example of what you may find when searching for collocations placed there, depending on the choice of word, is the one Teubert & Čermáková (2007:89–92) found when analysing a sample of 200 citations from the Bank of English¹⁵ of first *globalisation* and then *globalization*. They noted that with the former spelling, among the most frequent collocates they found words like *anti* and *against*, whereas with the latter spelling, those two words were not at all among the most frequent collocates, instead they found words like *market* and *investment*. This could be taken as a parallel to what I found in my previous work on *Rom* and *Cikán* (Elmerot 2016). Let me point out again, that statistics shouldn't be used to analyse how the words are used in their context, but as a tool to categorise certain words or expressions in a larger source material like the SYN series. "Determining the usage of lexical items and coping with it are essential to the methodology of corpus linguistics." as Teubert & Čermáková (2007, 82) put it.

¹⁵ "The full Corpus contains 4.5 billion words. The Bank of English™ is a subset of 650 million words from a carefully chosen selection of sources, to give a balanced and accurate reflection of English as it is used today." <https://collins.co.uk/page/The+Collins+Corpus?>

Baker et al. described in their 2008 article that it is the combination of statistics, corpus linguistics and discourse analysis that makes the researcher avoid missing important facts, and the inclusion that “aims to do justice” to the material, even though the method used here, due to the time-limit, “may miss or disregard strong non-adjacent collocates” (Baker et al. 2008, 275–276). Using this method means also to ignore a number of search hits that may be very relevant indeed for a socio-linguistic study or a qualitative discourse analysis to see the context and the actual themes in articles about these people. Adjectives further away from the keyword are excluded from this research, even if they could have been interesting in order to analyse the article’s theme in a context of Roma, Vietnamese or Ukrainians. Some researchers may view this as a methodical problem, but the aim here is to get a comprehensive picture that I believe is reached when the source material is this big, and the statistics give us a clearer picture.

A methodical problem that this study does have, is when positive or neutral adjectives (like *přizpůsobený* = adaptable) before the keywords are actually written with a negation, in this case *nepřizpůsobený*, which means inadaptable, a word with a much more negative connotation (cf. Slavíčková 2015). For some reason, the corpus filters do not (yet) distinguish between negated and non-negated adjectives, but instead put all words as non-negated in the frequency results. This means that there is some manual analysis necessary: the result was checked manually to see what adjectives were actually negated.

Adjectives that seem neutral or positive, but are in fact stereotypical, is another methodical problem. One example is the adjective *slušný*, decent, before *Rom*, since there is a stereotype about Roma not being decent (whatever definition one may use for that word). These kinds of adjectives are therefore trickier to categorise and sort, from a purely statistical viewpoint, and make the post-analysis work more time-consuming. Due to the time limits of this study, I’ve based the classification to a large extent on previous research on the minority groups in question, but will hopefully be able to do it more thoroughly in a future study.

The statistical processing of the material begins with sorting and picking out the adjectives that are evaluated as positive, negative and neutral of value, respectively. I'm not using the terms positive and negative here as in gradable adjectives (Sassoon 2010, 141–142), instead I use the terms as a measure of how the reader is likely to accept it, as something good and positive, or bad and negative, about the people in question. My main focus is therefore positive adjectives, negative adjectives and neutral adjectives, and for each group of people, I've added the stereotypical aspect to the adjectives. Since these groups are perceived differently, there are different stereotypes for each group, which creates a need to make the categorisation more detailed. For example, all three groups are, according to previous research, considered to behave violently and being the victims of violence, so adjectives like “beaten” or “attacked” were considered stereotypically negative, whereas any adjectives related to sports adjacent to Ukrainians were considered stereotypically positive. This latter categorisation is not based on previous research, since there was no mention of Ukrainians being very sporty there, but instead a result of the analysis for this study, when it turned out that so many adjectives adjacent to Ukrainians were related to sports events.

During the sorting, it became obvious that Roma are very often categorised by where they come from, so the geographical adjectives (as in “Czech” or “Living in Kolín”) were put in a neutral category (non-stereotypical neutral). Ukrainians and Vietnamese were on the other hand often categorised by their age, so there, the age adjectives were put in the same, non-stereotypical neutral, category.

7.1. Work steps

When I have logged in to the Czech National Corpus, I start with clicking the tab konText, let the choice “Basic” search be, since that also includes lemmata (in my cases, since all my search words are lexicological words), and add the context of having an adjective within one space to the left of the main keyword. I also check that the source language is Czech. When the hits are shown, I sort them on publishing date, just to get the hits that regard the same event close by each other – it helps when I

later need to sort out adjectives that are irrelevant for some reason (e.g. the name of a football club or other homonyms). After that, I go to the tab Frequency and do a custom sorting for frequency (“Multilevel frequency distribution”) of the lemma 1 step to the left of the keyword.

After exporting it all into three Excel files, I start classifying the adjectives manually, in one file each for the sentences including Rom, Vietnamec and Ukrajinec, sort them according to their respective category, and calculate the percentage sum for each category, to see how large a percentage each category covers. I have to take a second look at the context of the adjectives that may be negated, and note whether or not they are, I have to look at the context to make sure certain adjectives mean what they normally mean (like *vybraný*, that may mean both “representative” or “cultivated” and “looted” or “plundered”), and in the case of the Vietnamese I have to check whether or not the adjective comes from a text on the Vietnam war. In a few cases the adjectives adjacent to Ukrainians were from articles on the “orange revolution” there, something which may change the classification. In short, I needed to look out when classifying the adjectives I find – more so than in my previous study on Roma (Elmerot 2016, 17–20).

For this, time-limited, study, I’ve omitted the adjectives that had a frequency of less than one per mille (1 ‰), since they can hardly be called collocates. Of these, two homonyms were found: the football club AS Roma from Rome in Italy was mentioned with a frequency of 47, which is only 0.16 per cent of the total, and a Read-only Memory, abbreviated ROM, with the prefix “flash” had a frequency of 40, which is 0.13 per cent of the total. The adjective “full” occurred a few times, but only meant “full of” the people in question, so this was also removed from the result tables.

8. Results

8.1. The most frequent adjectives

To begin with, the twenty most frequent adjectives were sorted out. There were three searches, one for Rom, one for Ukrajinec and one for Vietnavec.

Table 1: The 20 most frequent adjectival collocations for each term (absolute frequencies per term).

	Rom (29,657)	Vietnavec (4,470)	Ukrajinec (5,335)
1	český (2,937)	mladý (380)	mladý (294)
2	mladý (2,217)	další (141)	bohatý (168)
3	slovenský (2,159)	český (128)	opilý (135)
4	místní (2,042)	jiný (100)	další (128)
5	olašský (1,582)	malý (92)	třicetiletý (97)
6	malý (704)	třiačtyřicetiletý (88)	mnohý (93)
7	rumunský (542)	zadržený (71)	zraněný (87)
8	další (540)	obžalovaný (60)	šestadvacetiletý (72)
9	ostravský (537)	místní (59)	pěťadvacetiletý (66)
10	vsetínský (453)	mrtvý (53)	čtyřadvacetiletý (65)
11	zdejší (449)	mnohý (52)	známý (63)
12	samotný (433)	čtyřicetiletý (51)	starý (60)
13	zaměstnaný (344)	samotný (48)	dvaatřicetiletý (59)
14	ostatní (300)	pětadvacetiletý (48)	pracující (59)
15	tamní (295)	ubitý (48)	dvaadvacetiletý (58)
16	mnohý (246)	známý (47)	devětadvacetiletý (57)
17	jiný (231)	zraněný (46)	zaměstnaný (56)

18	brněnský (226)	dvacetiletý (39)	sedmadvacetiletý (55)
19	přízpůsobivý (224)	zdejší (38)	třidvacetiletý (55)
20	starý (209)	sedmadvacetiletý (38)	osmadvacetiletý (55)
Share of total (per group)	56.2 %	36.4 %	33.4 %

From these twenty examples per group, it's easy to see that geography is more important for Roma than for Vietnamese or Ukrainians, and that there are fewer adjectives that are often used as collocations for Roma ("Czech" being ten per cent of the total in itself). Even if the articles often concern crimes and police reports, there are several mentionings of the "good" Vietnamese and Ukrainians, as if the interviewer or interviewee wants to emphasize that there are good and bad among these people. There are not as many examples of this when it comes to the Roma statistics. However, the idea that Ukrainians drink more than normally expected is shown by the fact that "drunk" (*opilý*) is the third most frequent adjective for Ukrainians. On the other hand, "working" (*pracující*) and "employed" (*zaměstnaný*, sometimes only meaning "working") are also among the 20 most frequent adjectives for Ukrainians, albeit on place 14 and 17, respectively.

The othering structures in previous research are confirmed by this study, although the amount of geographical (in the case of Roma) and age-related (in the case of Ukrainians and Vietnamese) adjectives was more abundant than expected. When these adjectives are ignored, there are mostly stereotypical negative, or pejorative, adjectives left, from articles on criminality (*zraněný*, "injured", *zadrženy* "detained", etc.) and drunkenness (*opilý*, "drunk", a few occurrences of *zfetovaný*, "stoned" in all three groups). However, I see nothing on the religiousness of Ukrainians, but I do see many occurrences of words relating to success, from *bohatý* to *favorizovaný*. I haven't found any mentions of this latter aspect of the Ukrainian immigrants in the previous research on the Czech Republic, unless that is what Leontiyeva (2004, 27) means with her duality on "guest working" and "non-problematic" Ukrainians.

But since the Roma have been living in the Czech lands for so long, it seems to be more interesting where they are from than anything else. The Ukrainians and Vietnamese are not as integrated in the Czech society, and therefore their collocates are much more often definitions of their age. However, the Vietnamese have *český* (“Czech”) as their third most frequent adjective, *místní* (“local”) as their ninth most frequent one, and *zdejší* (“local, resident”) as their ninth most frequent adjective, which is of course because these specific Vietnamese have been living in the country for long enough time. The Ukrainians are probably too much of a proximate other to ever be called “Czech”.

8.1.1. Adjectives before Roma

Compared to my previous study (Elmerot 2016), it seems clear that when a journalist writing in Czech chooses the noun *Rom* for a person instead of *Cikán*, the negative collocations and adjacent adjectives are comparatively few and far between. The infamous adjective *nepřizpůsobivý* – in context always written negated – has a frequency of 0.7 per cent before *Rom*, compared to 0.07 per cent before *Vietnamese* (and none before *Ukrainian*), but 1.23 per cent before *Cikán* in the previous study (Elmerot 2016, 18). Interestingly, the seemingly positive *zaměstnaný* (“employed” or just “working”) is in the Roma context negated, except when it’s in the negative context of “black”, illegal, labour (*načerno*). Among the neutral adjectives, there are two among the top ten that are not geographical (*malý* and *další*), and none is related to age.

8.1.2. Adjectives before Vietnamese

Often, the Vietnamese in the Czech republic are considered to keep to themselves, which is here represented by the word *samotný* (either “alone” or “him-/her-/themselves”) being so frequent. Words like “small” (*malý*) are here classified as stereotypical, due to the “infantilization” of Vietnamese that often occurs in Czech media (see [above](#), Alamgir 2013, 79). Since Vietnamese are also often mentioned primarily in articles regarding crime, words where they are clearly the culprit (e.g. *zadrženy, obžalovaný*) are classified as stereotypical – as long as they’re not taken from an article on the Vietnam war – whereas “dead” (*mrtvý*) is not, since that may refer to different kinds of death, and several from articles on war. Of the adjectives in the table, *známý* (“known”, “familiar”) is often found negated, *neznámý*, and that often in articles on crime. Still, this has been classified here as

purely neutral. The seemingly positive *zaměstnaný* is often negated or, similarly to Table 1 above, in the context of illegal labour – but not always, why it’s here classified as neutral.

8.1.3. Adjectives before Ukrainian

The most surprising was to find *bohatý*, “rich” having such a high frequency. This is in context partly due to the murders of “one of the richest Ukrainians” in 1996 and two “rich” Ukrainians in 1998, well reported mostly in *Mladá Fronta DNES*, as well as some news on an oligarch, mainly covered in *Hospodarské Noviny*. The well-known, international financial magazine *The Forbes* also had an article in October of 2014, available from their website, that concerns how the rich Ukrainians try to safeguard their money¹⁶. This adjective is in 0.1 per cent of the cases exchanged for the synonym *zámožný*.

That Ukrainians often succeed in sports events is obvious by the adjectives *slavný* (“famous”), *favorizovaný* (“favoured” or “fancied”) and *vítězný* (“winning”), which are high-frequent collocations in articles about sports and games. And finally: the adjective *zaměstnaný* is here more often used in a positive context, and not nearly as often negated as with Roma and Vietnamese.

8.2. The classification of the adjectives

The adjectives that have at least one per mille of the frequency in total were classified, and the stereotypical classes were based on previous research as well as what became evident from the result itself. These adjectives were sorted, according to their classification, and then a subtotal was calculated for both frequency and percentage, as well as a ratio to see clearer how the amount of negative adjectives and positive adjectives, stereotypical or not, were compared to each other. The comparative results are shown in the following tables.

Table 2: A comparison of the most frequent adjective in each category, per group.

	Negative adjectives		Neutral adjectives		Positive adjectives	
	Non-stereo-	Stereotypical	Non-	Stereotypical	Non-stereo-	Stereo-

¹⁶ <https://www.forbes.com/sites/kenrapoza/2014/10/10/rich-ukrainians-struggle-to-find-safe-havens-as-currency-market-flatlines/#96e31f311d64>

	typical neg.	neg.	stereo- typical neu.	neu.	typical pos.	typical pos.
Rom	zavražděný (0.2 %)	(ne)zaměstnaný (1.2 %)	český (9.9 %)	kočovní (0.3 %)	slušný (0.6 %)	vzdělaný (0.5 %)
Vietnamec	mrtvý (1.2 %)	zadržený (1.6 %)	mladý (8.5 %)	malý (2.0 %)	zaměstnaný (0.5 %)	bohatý (0.5 %)
Ukrajinec	mrtvý (0.8 %)	opilý (2.5 %)	mladý (5.5 %)	kontrolovaný (0.2 %)	pracující (1.1 %)	bohatý (3.1 %)

The word “young” (mladý) is in the top three for all three groups. Where the geographical adjectives have a high frequency for the term *Rom*, the age-related ones seems to be the most interesting about the terms *Vietnamec* and *Ukrajinec*, although the Vietnamese sometimes are noted as living in the Czech Republic. There are comparatively fewer stereotypical negative adjectives before Roma, which may be connected to the more neutral connotations to this term compared to *Cikán*. Ukrainians, on the other hand, are definitely stereotypically depicted, with 15 per cent of the total being stereotypical adjectives and 65 per cent non-stereotypical. Both Ukrainians and Vietnamese have a high rating of “rich” (*bohatý*), although it's sometimes in a negative context when adjacent to the Vietnamese, and not nearly as frequent as for the Ukrainians.

Table 3: Percentage of each classification, and total amount of adjectives used less than in one per mille. The ratio = the share of negative adjectives divided by the share of positive adjectives. The higher number, the more negativity.

	Negative adjectives		Neutral adjectives		Positive adjectives		Less than 1 %	Neg. total	Pos. total	Ratio neg./pos.
	Non-stereotypical neg.	Stereotypical neg.	Non-stereotypical neu.	Stereotypical neu.	Non-stereotypical pos.	Stereotypical pos.				
Rom	1.6 %	5.7 %	72 %	0.6	0.4	1.0 %	18.5 %	7.3	1.4	5.21
Vietnamec	12 %	7.2 %	64 %	2.9 %	2.5 %	3.0 %	8.0 %	19.2	5.5	3.49
Ukrajinec	1.8 %	11 %	56 %	0.2 %	7.7 %	4.3 %	19 %	11.02	12	0.92

9. Concluding discussion

The research questions,

1. To what extent are these people described in a stereotypical manner?
2. Are the “culturally” more distant Roma and Vietnamese more pejoratively othered than the Slavic Ukrainians?

are answered thus:

It is clear that these minority groups are often defined by adjectives that are negative, according to previous research, in that the Roma are unadaptable (place 19 in frequency), unemployed and victims of violence, whereas the Vietnamese are mentioned as being victims, on occasion unadaptable (place 235 in frequency) but more often tiny or just different. The Ukrainians are never unadaptable, but they are drunk and armed, which is in line with previous research. On the other hand, the Ukrainians are rich and famous or at least a celebrated sports person, and Vietnamese are tiny in stature. Both Vietnamese and Ukrainians are very often young in this source material, Vietnamese more so than Ukrainians.

It is also clear that the Ukrainians are more often mentioned in a positive light (altogether 12 per cent, compared to the 1.4 for Roma and 5.5 per cent for the Vietnamese). It also seems clear, compared to Elmerot 2016 – and perhaps also due to the “lip service” paid officially (Slavíčková & Zvagulis 2014, 153) – that the journalists using the word *Roma* are comparatively more respectful than if they use the word *Cikán*. The Vietnamese are actually treated almost as negatively as the Roma in this source material – the ratio is 5.21 per cent for the Roma and 3.49 for the Vietnamese, where the Ukrainians negative to positive ratio is only 0.92 per cent.

The Slavic Ukrainians are more often described in a stereotypical manner than the Roma or Vietnamese. A whole 15.5 per cent of the adjectives are stereotypical for the Ukrainians, but only 6–

7 per cent for the Roma and Vietnamese. This may be due to the fact that they are so often mentioned in sports, and that it may then be Ukrainians living outside of the Czech Republic.

Most of the results from previous research were thoroughly confirmed with this method. The religiousness of the Ukrainians mentioned in previous research is however not among these results. In the case of the Roma, they are normally depicted more negatively than the others in the language of the Czech press (a ratio of 5.21), but the Vietnamese (with their ratio of 3.49) are not far behind: short in height and often badly wounded and beaten or accused of something, and Ukrainians are drunk but rich, and more often mentioned as working than the other two groups (and having a ratio of 0.92). If this study had included the keyword *Cikán*, “Gypsy” for the Roma, the results would have been different and more negative adjectives would have been in their column or row in the tables. All three groups have the adjective “young” among the top three most frequent collocations, which perhaps is due to the bias of the articles – young people make more stupid things, and more often get in the paper.

If a study is based on a very large corpus, one is able to see how power relations are reflected when the media addresses their “ideal subject” (Fairclough 2001, 41), in this case the Czech general public. This study is based on millions of words, where I’ve picked out and sorted the ones that are actually placed next to each other, which means that the most frequent ones are possible collocations – “drunk Ukrainian”, for example. This method shows which adjectives are more frequent with the keywords, and what patterns there are in the printed media discourse. The result is aligned with a lot of previous research, but gives a deeper perspective and other ways of measuring the language that ordinary people read on a daily basis, when eyeing through their regular paper or magazine.

The result of this specific study may form a future basis for a more in-depth, intersectional study of expressions used about, for example, different workers in the Czech Republic (like typical female titles, perhaps gay people of both sexes, or mentions of poor people), to give a comprehensive picture of the general othering, not just of groups from different places. Perhaps that would show that different

groups of people are mentioned with similar words, and that would be a sort of proof of a systematic othering. Another option would be to analyse the expressions and collocations used during different periods of time, a diachronic study of othering in the printed Czech media, which seems not to have been done with the help of corpora yet.

References

- Alamgir, Alena K. 2013. "Race is elsewhere: state-socialist ideology and the racialisation of Vietnamese workers in Czechoslovakia". *Race & Class* Vol. 54(4):67–85.
- Baker, P.; Gabrielatos, C.; Khosravini, M.; Krzyzanowski, M.; McEnery, T. & Wodak, R. 2008. "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press" *Discourse and Society* 19:273–306.
- Bullock, A., Trombley, S. & Lawrie, A. (red.) 1999. *The new Fontana dictionary of modern thought*. 3. ed., fully rev. and updated. London: HarperCollins
- Cvrček, V.; Čermáková, A.; Křen, M. 2016. "Nová koncepce synchronních korpusů psané češtiny". *Slovo a Slovesnost: časopis pro otázky teorie a kultury jazyka*, 77(2):83–101. Praha: Ústav pro jazyk český, Akademie věd České republiky
- Doubek, D.; Levínská, M.; Bittnerová, D. 2015. "Roma as the Others". *Intercultural Education*, vol. 26(2):131–152.
- Elmerot, I. 2016. "Är en zigenare mer oanpassningsbar än en rom? En pilotstudie om kollokationer för orden Cikán och Rom i modern, tjeckisk tidningstext" *Slovo: journal of Slavic languages and literatures* 57:9–23. Uppsala: Institutionen för moderna språk, Uppsala universitet.
- Fairclough, Norman (2015). *Language and power*. 3., [updated] ed. London: Routledge
- Fidler, M. & Cvrček, V. 2015. "A data-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis". *Journal of Slavic Linguistics* 23(2):197–239.
- Fidler, M. 2016. "The others in the Czech Republic: Their image and their languages". *Multilingualism and Minorities in the Czech Sociolinguistic Space, a special issue of the International Journal of the Sociology of Language* 238:37–58.
- Hantonová, Barbora, 2013. "Mediální obraz Vietnamu v tištěných médiích". Master's thesis in Media and Communication, Faculty of Social sciences, Charles University, Prague.

- Karhanová, K., Homoláč, J., & Nekvapil, J. 2003. *Obraz Romů v střeoevropských masmédiích po roce 1989*. Brno: Doplněk.
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P., Zasina, A.: Corpus SYN, version 4 z 16. 9. 2016. Ústav Českého národního korpusu FF UK, Praha 2016. Available online: <http://www.korpus.cz>
- Leontiyeva, Y. 2004. "Pozadí výzkumu – přehled stavu zkoumaných komunit usazených v Česku, 5.A Ukrajinci." *Výzkumná zpráva: Integrace cizinců v ČR. Studie arménské, vietnamské a ukrajinské komunity v Praze a Středočeském kraji*. Praha: International Organization for Migration, p. 27.
- Leontiyeva, Y. 2006. *Menšinová problematika v ČR: komunitní život a reprezentace kolektivních zájmů: (Slováci, Ukrajinci, Vietnamci a Romové)*. Praha, Sociologický ústav AV ČR.
- Leontiyeva, Y. 2014. "The Education–Employment Mismatch among Ukrainian Migrants in the Czech Republic" *Central and Eastern European Migration Review* 3(1):63–84.
- Loomba, A. 2015. *Colonialism/postcolonialism*. Third edition. Milton Park, Abingdon, Oxon: Routledge
- Lüdeling, A. & Kytö, M. (red.) 2009. *Corpus linguistics: an international handbook. Vol. 2*. Berlin: Mouton de Gruyter
- McEnery, T & Hardie, A. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press
- Ottův slovník naučný: illustrovaná encyklopaedie obecných vědomostí. Pátý díl.* (1892). Praha: J. Otto. <https://ia802608.us.archive.org/35/items/ottvslovnknauni55ottogoog/ottvslovnknauni55ottogoog.pdf> [14 July 2016].
- Podolyan, I. & Nývltová, D. 2012. "Czech and Ukrainian Communication Patterns: A Cross-Cultural Research." *Zeitschrift für Slawistik*, 57(3):313–329.
- Said, Edward W. 1995. *Orientalism: [Western conceptions of the Orient]*, Repr. with a new afterword, Penguin, London.
- Sassoon, G. W. 2010. "The degree functions of negative adjectives". *Natural Language Semantics* 18:141–181.

Schneeweis, A. A. G. 2009. *Talking difference: discourses about the Gypsy / Roma in Europe since 1989*. (Diss.)

Slavičková, T. & Zvagulis, P. 2014, "Monitoring anti-minority rhetoric in the Czech print media: A critical discourse analysis". *Journal of Language & Politics* 13(1):152–170.

Slavičková, T. 2015. "Investigating nepřizpůsobivý as a key word in critical analysis of Czech press reports on Roma." *Korpus-Gramatika-Axiologie* 11:69–82.

Součková, T. 2015. "The Ukrainian Minority in Brno: A Qualitative Research on Ethnic Identity". *Ethnologia Actualis* 15(2):66–80.

Teubert, W. & Čermáková, A. 2007. *Corpus linguistics: a short introduction*. London: Continuum

Zalabáková, J. 2012. "Afričané, Romové a další „cizinci“ v české mediální realitě". *AntropoWebZin* 2:79–87.

Internet sources

Český statistický úřad (2012) *Cizinci v ČR_ Vietnamci u nás zakořenili* | ČSÚ

https://www.czso.cz/csu/czso/cizinci_v_cr_vietnamci_u_nas_zakorenili20120214 [18 April, 2017]

Idem (2014) *Analýza – Národnostní struktura obyvatel* pp. 4 & 25

<https://www.czso.cz/documents/10180/20551765/170223-14.pdf> [18 April, 2017]

Idem (2015) *T01 Cizinci celkem podle státního občanství k 31. 12. 2015*

https://www.czso.cz/documents/11292/27914491/1512_c01t01.pdf/7e9ea54d-13c8-4a41-bffa-bfe518d66ddc?version=1.0 [18 April, 2017]

Idem (2016) *T11 Cizinci podle typu pobytu, pohlaví a státního občanství - k 31. 12. 2016*

https://www.czso.cz/documents/11292/27914491/1612_c01t11.pdf [18 April, 2017]

Collins Learning *Collins for Education, Revision, ELT, Dictionaries & Atlases - The Collins Corpus*

<https://collins.co.uk/page/The+Collins+Corpus?> [18 April, 2017]

Merriam-Webster (2017) *Collocation* <https://www.merriam-webster.com/dictionary/collocation>

[18 April, 2017]

PARS International Corporation (2014) *Rich Ukrainians Struggle To Find Safe Havens As Currency, Market Flatlines* <https://www.forbes.com/sites/kenrapoza/2014/10/10/rich-ukrainians-struggle-to-find-safe-havens-as-currency-market-flatlines/#96e31f311d64> [7 April 2017]

Petr Lhotka (2010) *Romové v České republice po roce 1989* <http://skola.romea.cz/cs/historie/romove-v-ceske-republice-po-roce-1989/> [18 April, 2017]

Petr Lhotka (2010) *Romové v první Československé republice*

<http://skola.romea.cz/cs/historie/romove-v-prvni-ceskoslovenske-republice/> [18 April, 2017]

Romea o.p.s. (2010) *Historie | Skola Romea - vzdělávací cyklus nejen o romech*

<http://skola.romea.cz/cs/historie/> [18 April, 2017]

Idem (2017) *Patrik Banga: I'm for collecting ethnic data in all areas, but the majority can't tell who is Romani* <http://www.romea.cz/en/news/czech/patrik-banga-i-m-for-collecting-ethnic-data-in-all-areas-but-the-majority-can-t-tell-who-is-romani-2> [18 April, 2017]