

BEYOND PHILOLOGY

AN INTERNATIONAL JOURNAL
OF LINGUISTICS, LITERARY STUDIES
AND ENGLISH LANGUAGE TEACHING

19/1

LANGUAGE AND COMPUTERS

Edited by Anna Bączkowska and Joanna Redzimska

WYDAWNICTWO UNIwersYTETU GDAŃSKIEGO
GDAŃSK 2022

EDITOR-IN-CHIEF

Danuta Stanulewicz

DEPUTY EDITOR-IN-CHIEF

Magdalena Wawrzyniak-Śliwska

SECTION EDITORS

Olga Sokołowska (Linguistics)

Ludmiła Gruszewska-Blaim (Literary Studies, Culture)

Olga Kubińska (Translation)

Anna Bączkowska (Language Acquisition, Academic Teaching)

Olga Aleksandrowska (Education)

Jadwiga Węgrodzka (Reviews, Reports, Interviews)

ASSISTANT EDITORS

Maria Fengler

Karolina Janczukowicz

Marcin Fortuna

Karol Plichta

Michał Golubiewski

Joanna Redzimska

Dorota Góreczna

Małgorzata Smentek

Ewelina Gutowska-Kozielska

Joanna Stolarek

EDITORIAL ASSISTANT

Aleksandra Wisniewska

PROOFREADERS

Martin Blaszk

Jean Ward

Sarah Flamminio

Tadeusz Z. Wolański

Robert Urbański

The Editors and Proofreaders are all affiliated with the Institute of English and American Studies, University of Gdańsk, Poland.

COVER DESIGN

Andrzej Taranek

ISSN 1732-1220

eISSN 2451-1498

© Copyright by Uniwersytet Gdański

Wydawnictwo Uniwersytetu Gdańskiego

Contact address

Institute of English and American Studies

University of Gdańsk

ul. Wita Stwosza 51

80-308 Gdańsk

Poland

Phone: (+48) 58 523 30 49, (+48) 58 523 30 50

Email: beyond.philology@ug.edu.pl

ASSOCIATE EDITORIAL BOARD

Marta Bogdanowicz (University of Gdańsk, Poland)
Joanna Burzyńska-Sylwestrzak (Uczelnia Lingwistyczno-Techniczna,
Świecie, Poland)
Ewa Dąbrowska (Northumbria University, Newcastle, U.K.)
Desmond Graham (University of Newcastle, U.K.)
Gabriela Kelemen (Aurel Vlaicu University of Arad, Romania)
Zoltán Kövecses (Eötvös Loránd University, Budapest, Hungary)
Ronald W. Langacker (University of California at San Diego, U.S.A.)
Barbara Lewandowska-Tomaszczyk (University of Łódź, Poland)
David Malcolm (SWPS University, Warszawa, Poland)
Irene Gilssen Nordin (Dalarna University, Falun, Sweden)
Elżbieta H. Oleksy (University of Łódź, Poland)
Adam Pasicki (Pedagogical University of Kraków, Poland)
Piotr Ruszkiewicz (Pedagogical University of Kraków, Poland)
Otilia Costa e Sousa (Instituto Politécnico de Lisboa, Portugal)
Bogdan Szymanek (Catholic University of Lublin, Poland)
Ryszard Wenzel (Akademia Polonijna, Częstochowa, Poland)
Marta Wiszniowska (Nicolaus Copernicus University, Toruń, Poland)

BOARD OF REVIEWERS

Frank Cioffi (Baruch College, City University of New York, U.S.A.)
Anna Cisło (University of Wrocław, Poland)
Artur Czapiga (University of Rzeszów, Poland)
Tadeusz Danilewicz (Wyższa Szkoła Bankowa, Gdańsk, Poland)
Danuta Gabryś-Barker (University of Silesia, Katowice, Poland)
Aleksandra Kędzierska (Maria Curie-Skłodowska University, Lublin,
Poland)
Ewa Komorowska (University of Szczecin, Poland)
Wojciech Kubiński (Państwowa Wyższa Szkoła Zawodowa w Elblągu,
Poland)
Elżbieta Mańczak-Wohlfeld (Jagiellonian University, Kraków, Poland)
Marzenna Mioduszevska (Universidad Rey Juan Carlos, Madrid, Spain)
Grzegorz Moroz (University of Białystok, Poland)
Wojciech Nowicki (Państwowa Wyższa Szkoła Zawodowa im. Szymona
Szymonowica, Zamość, Poland)
Kazimierz Sroka (University of Gdańsk, Poland; Polonia University in
Częstochowa, Poland)
Krystyna Stamirowska (Jagiellonian University, Kraków, Poland)
Yuri Stulov (Minsk State Linguistic University, Belarus)
Kamila Turewicz (Państwowa Wyższa Szkoła Zawodowa w Płocku, Poland)
Agnieszka Uberman (University of Rzeszów, Poland)
Jacek Waliński (University of Łódź, Poland)
Tomasz Warchoń (Georgia Southern University, Statesboro, U.S.A.)
Julian Wolfreys (Independent Scholar, U.K.)

Beyond Philology is published in print and online:

<https://fil.ug.edu.pl/wydzial/instituty_i_katedry/institut_anglistyki_i_amerykanistyki/czasopismo_naukowe_beyond_philology>,
<<http://cwf.ug.edu.pl/ojs/index.php/beyond>>.

Licence: CC-BY

The online version is primary.

Beyond Philology is indexed by

- The Central European Journal of Social Sciences and Humanities
- ERIH PLUS European Reference Index for the Humanities and Social Sciences
- Index Copernicus
- MLA International Bibliography
- Norwegian Register for Scientific Journals, Series and Publishers

BEYOND PHILOLOGY 19/1

Contents

Foreword JOANNA REDZIMSKA, ANNA BĄCZKOWSKA	7
Infinitive-gerund alternations in Polish post-verbal complementation: A corpus-based study AGNIESZKA KALETA	9
Lexical factors in non-finite complementation of <i>continue</i> JOANNA PODHORODECKA	37
You shall know a term by the company it keeps: Collocations of the term <i>evidence</i> in general and legal corpora TOMASZ MICHTA	65
An attempt to reconstruct the linguistic picture of a true Pole IWONA BENENOWSKA	97
How much truth is there in a true Pole? IWONA BENENOWSKA	135
Forms of address and (im)politeness: A corpus-assisted study of Polish professional and non-professional subtitles ANNA BĄCZKOWSKA	159

Selected aspects of sentiment analysis in the context of human communication JOANNA REDZIMSKA	193
How concepts are understood – online collaborative work and digital data BARBARA LEWANDOWSKA-TOMASZCZYK	213
Using the LARA platform to crowdsource a multilingual, multimodal Little Prince ELHAM AKHLAGHI, ANNA BĄCZKOWSKA, BRANISLAV BĚDI, HAKEEM BEEDAR, CATHY CHUA, CATIA CUCCHIARINI, HANIEH HABIBI, IVANA HORVÁTOVÁ, CHRISTÈLE MAIZONNIAUX, NEASA NÍ CHIARÁIN, NATHALIE PATERSON, CHADI RAHEB, MANNY RAYNER, CHUNLIN YAO	245
Teaching sustainability communication in higher education: Applying case study method based on social media DALIA GULBINSKIENĖ, GIEDRĖ VALŪNAITĖ OLEŠKEVIČIENĖ	279
Information for Contributors	302

Foreword

JOANNA REDZIMSKA
ANNA BĄCZKOWSKA

The advent of high technology has influenced all walks of life, including research methods not only in science but also in humanities. Without doubt, research in linguistics (especially in corpus linguistics) has substantially profited from approaches and tools that facilitate analysing linguistic data, thus changing to a significant degree the research methodology in linguistics from a time-consuming and human-dependent one-to-one into performed faster and more software-oriented. As a result, it is common practice that linguists apply quantitative analyses where they cope with substantial numbers of language examples that can be browsed and grouped according to needed rules. However, software for processing natural languages or language corpora are also used for qualitative analyses when certain language tendencies or realizations are scrutinized with reference to detailed information, models or construals that they exemplify.

The selection of articles presented in the following part proves how informative and at the same time important the above methods are in connection to language analysis. As follows, the major group of articles focus on the corpus analysis of language data (Benenowska, Kaleta, Michta, Podhorecka, Lewandowska-Tomaszczyk, Akhlaghi et al., Bączkowska) where selected aspects of language issues are grouped and analysed from a particular perspective. Additionally, the articles by Gulbinskienė and Oleškevičienė, and Redzimska focus on selected aspects of language data that provide valuable insight into the nature of

their representation. Two texts revolve around the use of crowdsourcing methods in language teaching and translation (Lewandowska-Tomaszczyk, Akhlaghi et al.).

Finally, it is worth highlighting that a corpus analysis of language data along with its quantitative examination meet the challenges of the modern world. Apart from AI solutions that make use of natural language processing, language data are valuable for education or social policy making.

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

<https://doi.org/10.26881/bp.2022.1.01>

**Infinitive-gerund alternations in
Polish post-verbal complementation:
A corpus-based study**

AGNIESZKA KALETA

*Received 25.12.2021,
received in revised form 28.11.2022,
accepted 29.11.2022.*

Abstract

The paper constitutes a contribution to semantically based research on the so called ‘syntactic alternations’, defined as “structurally and/or lexically different ways to say functionally very similar things” (Gries 2017: 8). More specifically, the study is concerned with Polish verbal predicates which accept both the gerundive and infinitival complement, with the difference between these two alternatives being a matter of construal or focus rather than any clear-cut semantic contrasts. Taking as its point of departure the cognitive linguistic thesis that a change in form always entails a change in meaning, the present paper seeks to determine the factors that significantly affect the choice of one or the other complement type in situations where both are fully acceptable. The study provides an in-depth, corpus-based analysis of three selected pairs of constructions, and offers a new theory (referred to as *reduced-increased involvement theory*) to explain the choices that Polish speakers make with respect to the two constructions in question.

Keywords

complementation, gerund, infinitive, syntactic alternations

**Alternacje gerundium-bezokolicznik
w polskojęzycznych konstrukcjach
dopełnieniowych: Podejście korpusowe****Abstrakt**

Niniejszy artykuł wpisuje się w nurt badań nad semantyką tzw. ‘alternacji składniowych’, które można zdefiniować jako alternatywne sposoby wyrażania podobnych treści przy pomocy odmiennych środków składniowo-leksykalnych (Gries 2017: 8). Przeprowadzone badanie dotyczy polskojęzycznych czasowników, które współwystępują z dwoma różnymi formami dopełnieniowymi – gerundium (rzeczownikiem odczasownikowym) i bezokolicznikiem, tworząc tzw. pary minimalne, a więc konstrukcje odnoszące się do tej samej sytuacji, lecz różniące się pod względem sposobu jej konstruowania. Przyjmując za punkt wyjścia kognitywną tezę, że zmiana formy zawsze oznacza zmianę znaczenia, artykuł podejmuje próbę określenia czynników semantycznych wpływających istotnie na wybór dopełnienia gerundialnego lub bezokolicznikowego w sytuacjach gdy obydwie formy są równie akceptowalne. Artykuł przedstawia szczegółową analizę trzech wybranych par konstrukcyjnych przeprowadzoną w oparciu o metodologię korpusową, oraz proponuje nową teorię wyjaśniającą wybór formy dopełnieniowej przez natywnych użytkowników języka polskiego (nazwaną wstępnie „teorią ograniczonego-wzmożonego zaangażowania”).

Słowa kluczowe

dopełnienie, gerundium, bezokolicznik, alternacje składniowe

1. Introduction

Research on complement clauses has been thriving since the 1970s. While the early generative studies were preoccupied

mainly with the syntactic aspects of complementation, the more recent functional approaches have concentrated primarily on issues of semantics, offering many insightful observations about the relationship between the form and function of different types of complement clauses (cf. Bresnan 1979; Givón 1980, 2001; Noonan 1985; Wierzbicka 1988). The issue of syntax-semantics interface is also an important theme in complementation research inspired by the framework of construction grammar: a family of different, yet related theories, which define grammar as a vast, monostratal repository of *constructions*, i.e. learned pairings of form with meaning (cf. Goldberg 1995, 2006; Langacker 1987, 1991, 2008). Hence, when seen from the constructional perspective, different types of complement clauses represent constructions in their own right, which entails that their distribution is not random or arbitrary but semantically motivated. One way of approaching this matter by constructional grammarians has been by surveying all of the verbs that take a certain kind of complement clause and establishing whether the verbs in each set share any semantic features which might represent a unitary value of the different complement types (cf. Dirven 1989; Smith 2008; Taylor and Dirven 1991). Another research strategy has been to focus on the verbs that take two or more complement types, usually without any obvious differences in meaning. This type of constructions can be subsumed under the general heading of syntactic alternations, a topic of much interest to construction grammarians over the last decade or two (cf. Gries 2017; Gries and Stefanowitsch 2004; Kaleta 2014; Rudanko 2021; Stefanowitsch and Gries 2003). The present study constitutes a contribution to this latter strand of research by focusing on alternations between the gerund and the infinitive, which can be found in post-verbal complement clauses in present day Polish. While there is ample literature on infinitive-gerund alternations in English complementation (cf. Deshores 2015; Deshores and Gries 2016; Kaleta 2012; Rudanko 2017, 2021; Smith and Escobedo 2001), the corresponding research on data from other languages is very scarce or even

non-existent, the result of which is that our hitherto knowledge on this type of alternations is heavily biased by the patterns found in the English language. The present paper aims to fill this gap by providing an analysis of complement patterns of selected Polish verbs that accept both the infinitival and the gerundive complement, as illustrated with the examples below:¹

- (1a) Polacy chcą wyjść z UE.
Lit. 'Poles want to leave EU.'
- (1b) Polacy chcą wyjścia z UE
Lit. 'Poles want leaving EU.'
- (2a) Zaczął pisać powieści w wieku 10 lat.
Lit. 'He started to write novels at the age of 10.'
- (2b) Zaczął pisanie powieści w wieku 10 lat.
Lit. 'He started writing novels at the age of 10.'
- (3a) Zakazali nam chodzenia do kościoła.
Lit. 'They forbade us going to church.'
- (3b) Zakazali nam chodzić do kościoła.
Lit. 'They forbade us to go to church.'

While it is true to say that in certain cases (such as the ones above) the two complement patterns can be used more or less interchangeably, there are numerous instances in which the substitution of one form for another produces clear-cut differences in meaning. Consider (4a) and (4b):

¹ I use the term 'gerund' to refer to nominalizations also known as *substantiva verbalia* – in Polish they are derived with a fully productive *-nie/-cie* suffix (e.g. *pisanie*, *mycie*). *Substantiva verbalia* should be distinguished from *substantiva deverbalia* – while the former are morphologically regular and can be formed from nearly all verbs, the latter are much more idiosyncratic and can be derived only from a very limited number of verbs (cf. Krzeszowki 1990: 208). Krzeszowki describes *substantiva verbalia* (e.g. *pisanie*, *mówienie*) as being 'more nouny' than *deverbalia* (e.g., *pisarstwo*, *mowa*) and compares them to Lees' Gerundive Nominals.

(4a) Polacy chcą oddzielenia edukacji od religii.
Lit. 'Poles want separating education from religion.'

(4b) Polacy chcą oddzielić edukację od religii.
Lit. 'Poles want to separate education from religion.'

Also, there are cases where the substitution of one form with another produces clearly unacceptable expressions, as illustrated in (5b):

(5a) *Zaczęło się odliczanie do Świąt Bożego Narodzenia.*
Lit. 'The counting down to Christmas has begun.'

(5b) **Zaczęło się odliczać do Świąt Bożego Narodzenia.*
Lit. 'To count down to Christmas has begun.'

The existence of such restrictions clearly indicates that the two complement patterns under consideration are semantically distinct, i.e. each of them makes its own semantic contribution to the meaning of the whole utterance. Hence, the main question that guides the present research is the following: what is the semantic import of the infinitival and gerundive complement? Differently put – what are the factors that motivate the choice of one complement type over the other in cases where both are accepted by a given matrix predicate? This question is addressed with corpus-based methods, which are used to gain an insight into the distributional patterns of the two complement constructions in question. More specifically, the paper presents three case studies that explore in some detail the contextual linguistic factors that influence the choice of one or the other complement type by three semantically distinct verbs: a verb of volition – *chcieć* (want), a verb of aspect – *zaczynać* (begin), and a negative directive – *zakazać* (forbid) (cf. examples 1-3). Underlying this research is Harris's (1954) *distributional hypothesis*, which presumes the existence of correlations between distributional characteristics of linguistic structures and their functions/meanings.

The paper is structured as follows. Section (2) outlines the details of the methodological procedure followed in this study. Sections (3)-(5) present the results of the three case studies, i.e. they determine and discuss the distributional features that significantly distinguish the infinitival constructions from the corresponding gerundive constructions. Section (6), on the other hand, provides a qualitative interpretation of the distributional/frequency data presented in the previous sections and offers a theory called *reduced-increased involvement theory* to explain the contrasts between the two constructions under investigation. More specifically, it is argued that the infinitive-gerund alternations encode different degrees of the matrix clause agent's involvement in the event described in the complement clause. It is shown that the gerundive constructions present the matrix agent(s) as being 'less involved' or 'less salient' and hence less in control over the complement event than the agents of the infinitival construction, who exert more control or are more directly affected by the action described in the complement clause. The last section concludes the discussion and offers some prospects for future research.

2. Methodological procedure

As pointed out in the introductory section, the study is situated within the framework of functional-cognitive grammar, which posits that all units of language, including syntactic forms such as the infinitive or gerund, are constructions in their own right, i.e. distinct form-meaning pairings, which means that their distribution is semantically motivated. This also entails that there are no fully synonymous constructions as "a difference in form always spells a difference in meaning" (Bolinger 1968: 127). Taking this assumption as a point of departure, the present study uses corpus data and corpus-based methods to argue for the semantic distinctiveness of the two nearly-synonymous complement constructions, as introduced in section (1): [verb + the infinitive] and [verb + the gerund]. The methodological

procedure consists of two steps. First, all the verbs taking the infinitival complement and the verbs complemented by the gerund have been extracted from the corpus. The corpus used for this study is the Polish Web 2012 corpus, available via Sketch Engine software.² Since this is a very large database consisting of circ. 812,818,518 words, a smaller sub-corpus of 43,226,158 was created in order to make the data more manageable and available for hand-editing if required. All the relevant patterns were extracted using the CQL (corpus query language) function of the Sketch Engine. Table 1 presents the codes used to extract the constructions in question, together with the corresponding token and type frequencies.

Table 1

Codes used for data extraction from Polish Web 2012

construction	CQL code	Tokens	Types
Verb+infinitive	[tag="V.*"][]{}0,2[tag="inf.*"]	463312	105
Verb+gerund	[tag="V.*"][]{}0,2[tag="ger.*"]	16712	191

In the next step, the verbs appearing with both the infinitival and gerundive complement were determined using a CQL code. Table 2 presents the verbs that occur more often with the gerund than the infinitive, while Table 3 lists the predicates that favour the infinitival complement over the gerundive one. The tables give the raw frequencies of each of the two constructions, as they occur with a given matrix verb.

² <https://www.sketchengine.eu/>

Table 2

Verbs with the higher incidence of gerundive complements

lemma	Inf.	Ger.
umożliwić/umożliwiać (enable)	5	762
rozpocząć/rozpocząć (begin)	55	379
zaprześć/zaprześć (stop/cease)	48	214
zakazywać/zakazywać (prohibit)	65	191
zalecać/zalecać (advise/suggest)	23	104
zlecić/zlecać (order/instruct)	2	61
sugerować (suggest)	26	74
życzyć (wish)	39	40

Table 3

Lemmas with the higher incidence of infinitival complements

lemma	Inf.	Ger.
chcieć (want)	51580	190
zacząć/zaczynać (begin)	29982	123
próbować/spróbować (try)	11723	29
woleć (prefer)	2884	21
lubić (like)	5015	69
pomóc/pomagać (help)	4140	94
pragnąć (desire)	2383	47
uczyć/nauczyć (teach)	2778	140
uwielbiać (love)	1055	21
planować/zaplanować (plan)	820	214
proponować/zaproponować (propose)	750	301
nakazać/nakazywać (order/require)	508	139
polecieć/polecać (tell/order)	452	148
skończyć/kończyć (finish)	370	120
obietcać/obietcywać (promise)	359	58
zabronić/zabraniać (forbid)	357	216
kochać (kochać)	221	21

Given the space limitations of the present paper, only three predicates have been selected for a detailed consideration, each representing a different semantic type: *chcieć*, *zaczynać* and *zakazać*. The analytical procedure at this stage also consisted of two steps. First, a sample concordance of each verb has been extracted from the sub-corpus and subjected to a preliminary qualitative inspection. Next, the concordance lines have been coded for the features that could potentially be significant in distinguishing the gerundive and the infinitival construction. These data were then subjected to the chi-square test in order to determine whether there is an association between a given variable and the constructional choices. Once the significant associations have been determined, the frequency data have been subjected to qualitative interpretation, on the basis of which a hypothesis concerning the semantic import of the constructions in question has been formulated. As pointed out in the introductory section, this approach is based on the distributional hypothesis, according to which the linguistic context in which a word occurs determines its meaning. Given that cognitive-functional models of grammar regard all units of language, irrespective of their degree of morpho-syntactic complexity, as form-meaning pairing, the distributional hypothesis can be easily extended to cover syntactic units. Hence, the assumption underlying this research is that the distributional properties of a construction are indicative of its semantic properties, which is to say that linguistic/contextual features provide the 'clues' to constructional semantics.

3. Case study 1: *Chcieć* and its complements

Even a cursory look at the relevant concordance lines makes it clear that the complement choices made by this basic verb of volition are significantly influenced by whether the subjects of the main clause and complement clause are coreferent or non-coreferent (i.e. whether they refer to the same individuals or to different individuals). As it appears, in the former case the

infinitive is selected, whereas in the latter case the gerundive complement is the preferred choice. Consider the following example:

(6) Chcemy całkowitego *oddzielenia* edukacji od religii.

Lit. 'We want the complete separation of education from religion.'

In this example the ones who want the separation of education from religion, most likely, are not the ones directly responsible for or actually carrying out the complement process. Note that the use of the infinitival complement instead of the gerundive one produces a clear shift in meaning in that in this case the subject of the matrix clause is clearly co-referent with the complement clause agent(s), as illustrated in (7):

(7) Chcemy całkowicie *oddzielić* edukację od religii.

Lit. 'We want to completely separate education from religion.'

The impact of the subject (non)coreferentiality on the complement choices of *chcieć* becomes even more visible in cases where the complement agent is specified rather than implicit. Compare (8) and (9):

(8) Czego chcemy od rządu? *Podniesienia* zarobków i ograniczenia wzrostu cen.

Lit. 'What do we want from the government? Raising the earnings and limiting the price increase.'

(9) Czego chcemy od rządu? **Podnieść* zarobki i ograniczyć wzrost cen.

Lit. 'What do we want from the government? To raise the earnings and limit the price increase.'

The above examples show that the gerundive construction is selected in situations when the main clause agent(s) is/are non-coreferent with the complement agent(s), with the latter typically remaining unspecified and not expressed linguistically. How-

ever, a closer look at the relevant usage patterns reveals that (non)coreferentiality is not an all-or-nothing affair. Consider the following examples:

(10) Putin *chce utworzenia* strefy wolnego handlu między UE a Rosją.

Lit. 'Putin wants creating a free-trade area between EU and Russia.'

(11) Turcja *chce wejścia* do Unii Europejskiej i akceptuje intensywne nadzór i liczne pytania o jej praktyki wewnętrzne.

Lit. 'Turkey wants entering the EU and accepts the intense supervision and numerous questions about its internal policies.'

In both these examples the matrix subject can be regarded as co-referent with the complement agent. Yet, this coreferentiality is only partial. That is, in (10) Putin, in most likelihood, is not going to be directly or personally involved in the process of establishing a free-trade zone. His role is more that of a controller of supervisor of the process being described than its direct participant. On the other hand, in (11) Turkey is not an independent or autonomous decision-maker regarding the process of EU enlargement – all member countries participate in decision making. What seems to follow from these examples is that (non)coreferentiality should be seen as a gradable category, with different degrees of the matrix subject's involvement in the complement process. That is, apart from strictly non co-referent uses, the gerundive complement can also be found in constructions with coreferent agents. Yet, in this latter case, the matrix subject's involvement in the complement process is subject to some restrictions or limitations in the sense that he/she is construed as one of the many agents responsible for the complement process. Note that in such situations the gerund can be replaced with the infinitive. Hence, one could also say:

(12) Putin *chce utworzyć* strefę wolnego handlu między UE a Rosją.
Lit. 'Putin wants to create a free trade zone.'

(13) Turcja *chce wejść* do Unii Europejskiej [...]
Lit. 'Turkey wants to enter EU [...]'

The difference between (10) and (12), on the one hand, and (11) and (13), on the other, is a subtle one, i.e. in the infinitival construction the matrix agent appears to be construed as the main effector of the complement process and therefore as being more in control of the process described in the complement clause than the subject of the corresponding gerundive constructions, whose involvement in the complement event is subject to some limitations, as argued above. The conclusion that follows from these considerations is that the (non)coreferentiality of the matrix and complement agents is a key factor in distinguishing between the gerundive and infinitival constructions, as they occur with *chcieć*. However, the distinction is a matter of degree rather than a clear-cut dichotomy. In order to substantiate this hypothesis with quantitative data, a corpus-based analysis has been performed on random samples of the two constructions under consideration, as extracted from Polish Web 2012 corpus (cf. section 2). All the occurrences have been coded for the variable of subject coreferentiality. Apart from the unambiguously coreferential and non-coreferential uses, the category of partial coreferentiality, as discussed above, has been distinguished. Also, some ambiguous cases have been found in the data, which could be hardly judged as being coreferent or non-coreferent from the limited amount of the context provided by the concordancer. The results of this analysis are shown below:

$$\chi^2_{\text{Pearson}}(3) = 100.64, p = 1.13\text{e-}21, \hat{V}_{\text{Cramer}} = 0.86, \text{CI}_{95\%} [0.70, 1.00], n_{\text{obs}} = 133$$

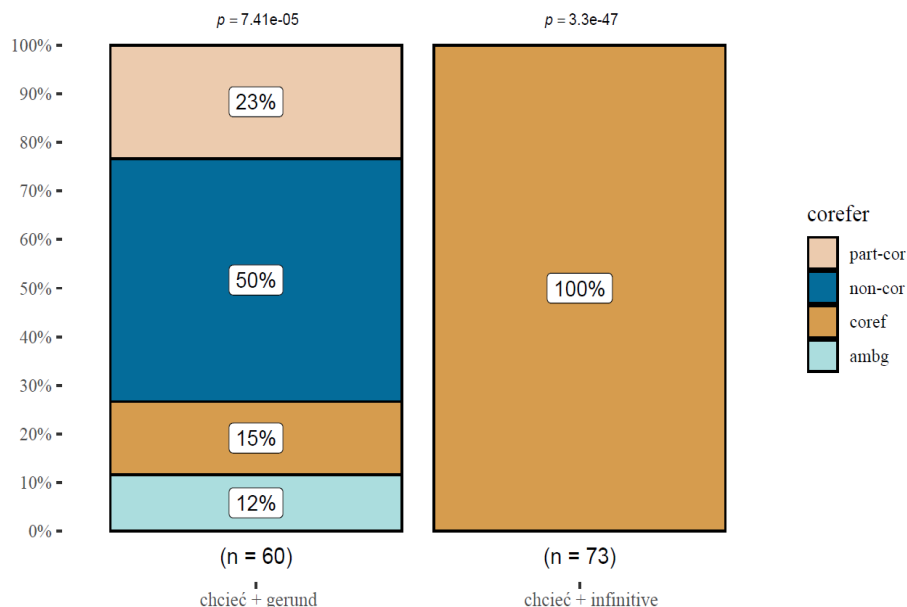


Figure 1

Coreferentiality of the subjects and the choice of the complement type by *chcieć*

The analysis has confirmed the significance of the coreferentiality variable: while the infinitival complement is selected exclusively by unambiguously coreferent constructions, the gerundive construction clearly favours non-coreferent uses, with unambiguous cases accounting for 50 % and partially coreferent ones for 23 % of the sample being analyzed. Note that coreferent uses constitute merely 15 % of the gerundive sample. As expected, the Person's chi-squared test confirmed that these differences are highly significant: $\chi^2(3) = 100.64, p = 1.13\text{e-}21$. The overall effect is large: Cramer's $V = 0.86$.

4. Case study 2: *Zacząć* and its complements

As can be seen from Tables 1 and 2, there are three verbs of aspect that accept both types of non-finite complements: *zacząć*, *rozpocząć* and *zaprzestać*. The two nearly synonymous verbs of inception – *zacząć* and *rozpocząć* differ in their preferred complement types in that the former occurs more often with the infinitive, whereas the latter preferentially selects the gerundive complement. It is interesting to note at this juncture that *rozpoczynać* is semantically distinct from *zacząć* in that it denotes the onset of iterative, generic processes or events, whereas *zacząć* is used primarily in the context of specific, one-shot actions performed by a definite agent. Hence, while (14) below is a perfectly well-structured sentence, (15) is rather awkward, as ‘*rozpocząć pisanie*’ suggests a long term process rather than an instantaneous action. Hence, (16) is a more natural alternative to (15).

(14) Japończycy *rozpoczęli* bombardowanie strategicznych rejonów Singapuru.

Lit. ‘The Japanese started bombarding the strategic regions of Singapore.’

(15) ?Wzięła do ręki długopis i *rozpoczęła* pisanie.

Lit. ‘She picked up a pen and started writing.’

(16) Wzięła do ręki długopis i *zaczęła* pisać.

Lit. ‘She picked up a pen and started to write.’

Also, it is worth pointing out here that *zacząć* is less restrictive and can be used as a (less formal) substitute for *rozpocząć*. Consider (17):

(17) Japończycy *zaczęli* bombardowanie strategicznych rejonów Singapuru.

Lit. ‘The Japanese began bombarding the strategic regions of Singapore.’

This distribution appears to suggest that the gerundive complement tends to form a tighter relationship with the predicates that favour iterative, repetitive processes. A further confirmation of this hypothesis might be the fact that the gerund is clearly incompatible with bodily actions performed by specific agents on specific occasions, which can hardly be conceptualized in iterative terms. In such cases, the infinitive is the only possible choice:

(18) Kiedy usłyszał to, *zaczął płkanie/mówienie/ubieranie się [...] Lit. 'When he heard this, he started crying/talking/getting dressed.'

(19) Kiedy usłyszał to, zaczął płakać/mówić/ubierać się ... Lit. 'When he heard this, he started to cry/talk/get dressed [...].'

Again, in order to verify this preliminary hypothesis a random sample of both constructions has been coded for the situation type denoted by the complement clause. This variable has been broken into three levels: single action, iterative action/ event, non-agentive process. The results are presented in Figure 2.

$$\chi^2_{\text{Pearson}}(2) = 20.13, p = 4.25\text{e-}05, \widehat{V}_{\text{Cramer}} = 0.32, \text{CI}_{95\%} [0.17, 1.00], n_{\text{obs}} = 182$$

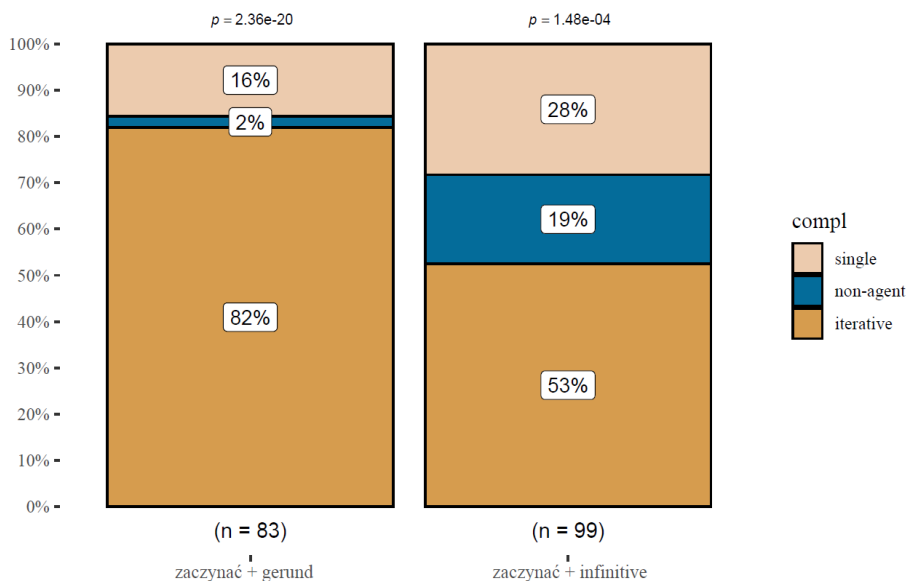


Figure 2

Complementation of *zaczynać* and subject types

As can be seen, both constructions are most often used to express iterative events. However, the gerundive complement is clearly skewed towards iterative actions, whereas the infinitival construction shows a more even distribution of the different situation types. Note also that non-agentive processes tend to be coded by the infinitival complement rather than the gerundive, which is illustrated with the following examples:

(20) Jesienią zaczyna robić się spokojniej.

Lit. 'In autumn it starts to get calmer.'

(21) (...) 32 proc. Brytyjek poniżej 30 roku życia zaczyna siwieć.

Lit. '32 per cent of British women under 30 starts to get grey.'

The Pearson's chi-squared test shows that there is a significant association between the type of situation coded by the comple-

ment clause and the type of the construction ($\chi^2(2, N = 182) = 20.1$; $p < 0,001$). As indicated by a post-hoc test applied to these data, the significant result is due to two types of association, i.e. between iterative vs. non-agentive events ($p_{adj} < 0.001$; $V = 0.34$), on the one hand, and iterative vs. single events ($p_{adj} = 0.015$; $V = 0.22$), on the other. At the same time no significant relationship has been observed between iterative vs. single events ($p_{adj} = 0.106$; $V = 0.25$). However, the overall effect is only moderate ($V = 0.33$).

Another variable that has been subjected to an empirical test is the type of the subject. The qualitative inspection of the concordance lines has revealed a strong presence of impersonal subjects in the gerundive sample. In particular, it has been observed that the gerundive complement tends to co-occur with the “się” impersonal reflexive forms, as illustrated below. Note that the infinitive is unacceptable in this context:

(22) W końcu zaczęło się odcinanie kuponów, zresztą bardzo dochodowe.

Lit. ‘Finally, clipping off coupons began, [...]’

(23) *Zaczęło się odcinać kupony, zresztą bardzo dochodowe.

Lit. ‘Coupons began to be clipped off [...]’

(24) Tak zaczęły się aresztowania i skazania winnych bratobójczej śmierci.

Lit. ‘It was how the arrests and convictions of the guilty of fratricidal death began.’

(25) *Tak zaczęły się aresztować i skazać winnych bratobójczej śmierci.

Lit. ‘It was how to arrest and convict the guilty of fratricidal death began.’

It is interesting to note here that the substitution of personal subjects for the impersonal ones makes the infinitive fully acceptable.

(26) W końcu zaczęła odcinać kupony.

Lit. 'Finally, she started to clip off the coupons.'

(27) Tak zaczęli aresztować i skazywać winnych bratobójczej śmierci.

Lit. 'It was how they started to arrest and convict the guilty of fratricidal death.'

Given these preliminary observations, the same sample as before has been coded for the subject type variable, which has been broken into two levels: impersonal constructions (i.e. with unexpressed agents) vs. personal constructions (i.e. with expressed agents). The results are presented in Figure 3:

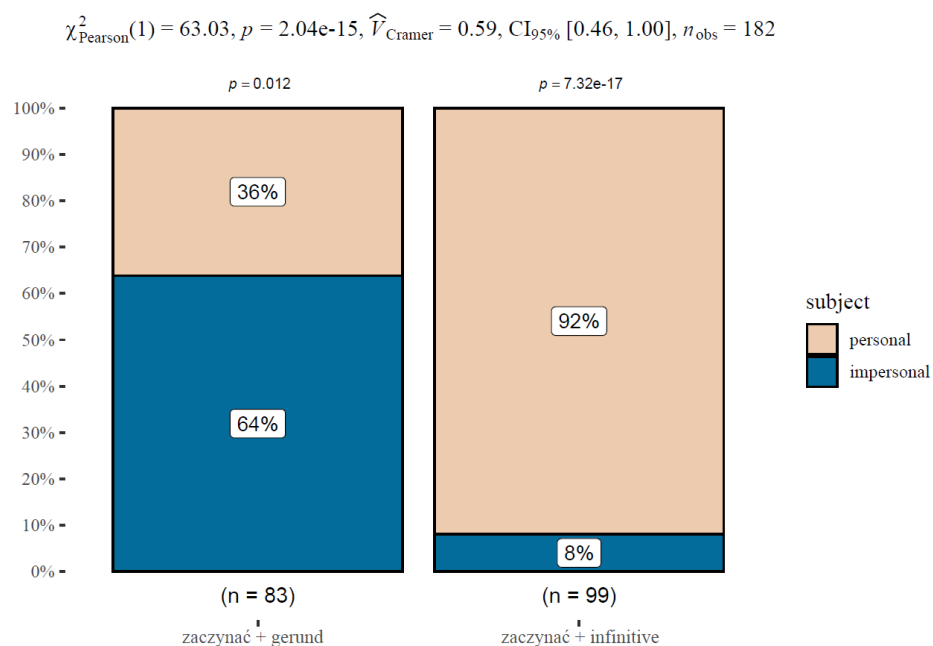


Figure 3

Personal vs. impersonal subjects in *zaczynać* constructions

As can be seen, the infinitival construction clearly favours personal subjects (92 %), with impersonal forms accounting for merely 8 % of the sample. In the case of the gerundive constructions, these proportions are reversed – the impersonal uses clearly dominate, accounting for 65 % of the sample. The chi-squared test confirmed that there is a significant association between the subject type and the construction ($\chi^2(1, N = 182) = 60.55; p < 0.001$). The overall effect is large $\phi = 0,59$.

To sum up, although the two constructions under consideration overlap to some extent, some clear preferences (and dis-preferences) can be observed in their usage patterns in terms of the type of the situation coded by the complement clause and the type of the subject referent in the matrix clause. In both cases there is a statistically highly significant correlation between the constructional choices and the variables in question. However, as indicated by the respective Cramer's V values, the association is much stronger in the case of the subject type than the situation-type variable.

5. Case study 3: *Zakazać* (forbid) and its complements

The notion of (im)personality also appears to play a significant role in the selection of the complement type by *zakazać*. Consideration of the relevant concordance lines has revealed that the infinitive is selected more often if the main clause subject is specific and also when the lower clause agent is given in the sentence and has a specific referent. The gerundive complement, on the other hand, favours generic subjects in both clauses and is also selected when the agents of the complement clause are unspecified and unexpressed. The following examples illustrate these two main patterns.

- (28) Ja mu nie zakazuję mówić w ojczystym języku [...].
Lit. 'I don't forbid him to speak his native tongue.'

- (29) W USA zakazano kiedyś picia alkoholu, a pił prawie każdy [...] Lit. 'In the USA drinking alcohol was once forbidden, yet everyone was drinking.'

As in the two previous studies, these preliminary observations have been subjected to quantitative verification. That is, a sample of the two constructions has been coded for the type of matrix subject and the type of the complement subject. The former has been broken into three levels: impersonal, generic, specific, and the latter into: not expressed, generic and specific. Figure 4 presents the distribution of the matrix subjects.

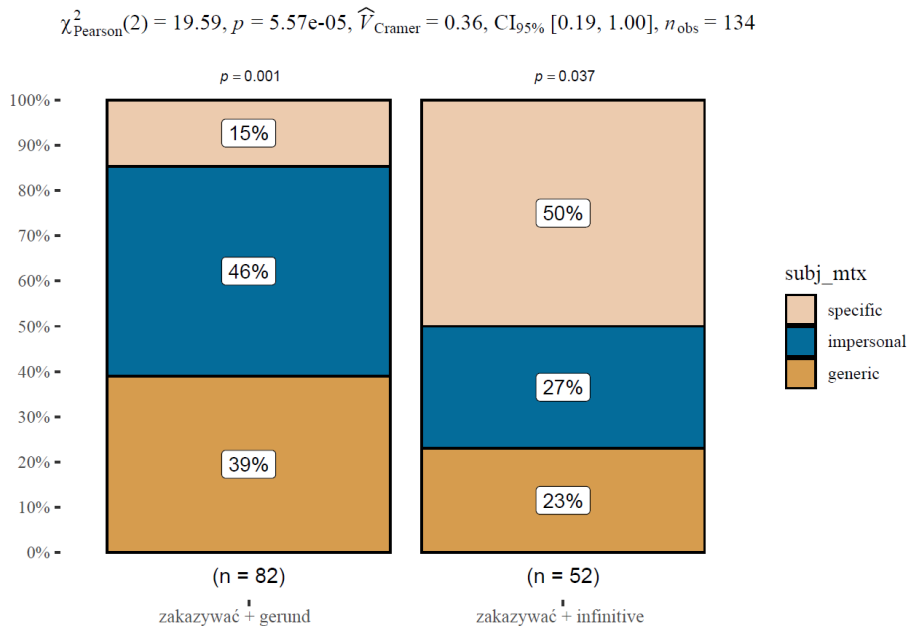


Figure 4

Type of the matrix subject in *zakazywać* constructions

As these frequencies indicate, the gerundive construction clearly favours impersonal and generic subjects in the main clause, which account for 46 % and 39 % of the sample, respectively,

with specific subjects being rather marginal (15 %). The infinitival construction, on the other hand, prioritizes subjects with specific referents (50 %) over the two other types. The chi squared test shows that the type of the matrix clause subject is highly significantly correlated with the choice of the complement construction ($\chi^2(2, N = 134) = 19.59$; $p < 0.001$). However, the overall effect is moderate $\eta^2 = 0.36$.

Let us now consider the type of the subject in the complement clause. Also here three levels have been distinguished: subjects with specific (definite) referent, subjects with generic referents and subjects that are not coded. The quantitative analysis has revealed the distribution presented in Figure 5.

$$\chi^2_{\text{Pearson}}(2) = 19.15, p = 6.94e-05, \widehat{V}_{\text{Cramer}} = 0.36, \text{CI}_{95\%} [0.19, 1.00], n_{\text{obs}} = 134$$

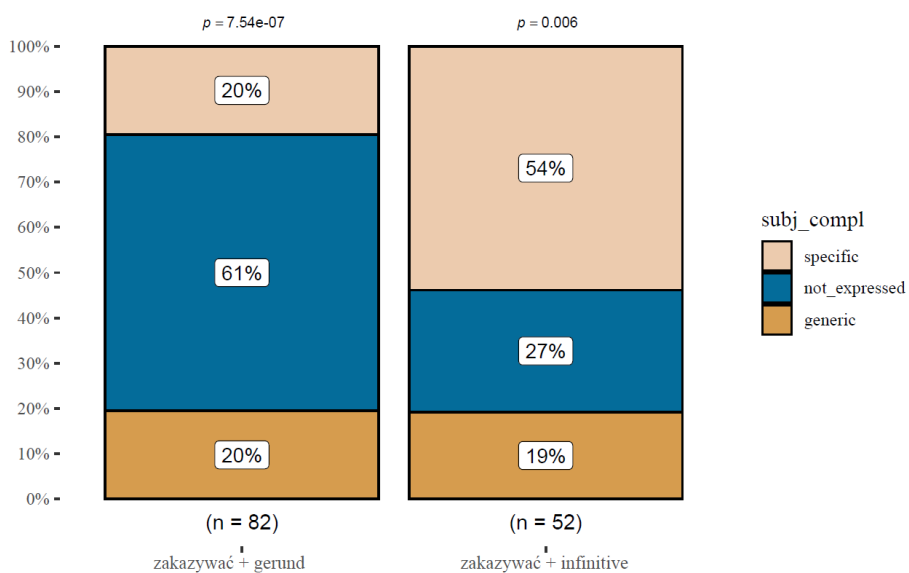


Figure 5

Type of complement subjects in *zakazywać* constructions

As can be seen, the distributional patterns of complement subjects are similar to the ones observed in the main clause: in the infinitival construction, specific subjects prevail (59 %) over the

two other types, whereas the gerundive construction clearly favours impersonal/non-expressed subjects (61 %). The chi-squared test confirmed that there is a highly significant correlation between the choice of the complement construction and the type of complement subject ($\chi^2(2, N = 134) = 19.1$; $p < 0.001$). The overall effect size is moderate $r\phi = 0.36$.

In sum, the obtained data allow to conclude that *zakazać* + gerund significantly favours impersonal or generic contexts of use, whereas *zakazać* + infinitive is significantly skewed towards specific situations, i.e. actions performed by specific agents. However, given that the correlation between the two variables is only moderate, some overlaps between the two constructions can be expected.

6. Reduced-increased involvement and dispersal of control

The three pairs of constructions considered in this paper denote distinct kinds of processes, which is hardly surprising given the semantic distinctiveness of the matrix predicates. Yet, the distributional data presented in the three case studies show a striking consistency in the way the two complement constructions under consideration portray the relationship between the main clause agent and the complement process. The pattern that emerges from the data is that the choice of one or the other complement type has to do with the degree of the matrix agent's control over the process described in the complement clause. As has been seen, the gerundive construction significantly prefers non-coreferent uses (cf. *chcieć*) as well as unspecified or generic subjects (cf. *zacząć*, *zakazać*) over the coreferent and definite/specific subjects. It should be clear that in non-coreferential constructions it is the complement subject, not the matrix subject that is the 'doer' of the complement action. Hence, the latter is only indirectly involved in the complement process and therefore has only limited influence on its outcome. It follows from this that the gerundive construction with *chcieć* profiles

the notion of reduced or limited control that the main clause agent has over the complement process, whereas the corresponding infinitival construction portrays the matrix agent as the main doer and hence the one having the complement action directly under his/-her control.

Also, we have seen that the notion of control is a matter of degree rather than being an all or nothing phenomenon in that the subject referent can be construed as one of the agents responsible for the complement process (cf. examples 10, 11). This situation involves what might be described as a 'dispersal or control' in the sense that the responsibility for the successful completion of the action is spread over different actors, rather than being exclusively in the hands of the matrix subject referent. As we have seen, in such intermediate cases *chcieć* can be used with both types of complements more or less interchangeably. Yet, the choice of the infinitive or the gerund involves a subtle shift in perspective, which proceeds in the direction predicted by this analysis, i.e. the infinitival constructions present the subject as being more in control, whereas the gerundive ones profile the notion of reduced controllability on the part of the matrix agent.

The type of subject has also been found to be a significant factor in the complement choices of *zaczynać* and *zakazywać*. These choices are significantly influenced by whether the agents are specified or unspecified. As has been seen, the gerundive constructions clearly favour unspecified/impersonal subjects, while the infinitival ones prefer definite/specified agents. This distribution appears to confirm the control/ involvement theory as formulated above in that the unspecified agents can hardly be construed as having any control or being actively involved into a process. In other words, if the agent is unknown, or unexpressed, his/her causal influence is either irrelevant or defocused. If, on the other hand, the agents are defined and specific, the issue of control becomes much more relevant and figures more prominently in the description of the scene.

It is advisable to point out here that *zakazać* is distinct from the two other predicates in that it obligatorily requires that the subject of the main clause is non co-referent with the subject of the complement clause. This, in turn, appears to explain why *zakazać* is more significantly associated with the gerundive complement, while the two other predicates favour the infinitival complement. Given that non-coreferentiality entails that the matrix subject is less in control or less directly involved in the complement action (for the simple reason that we have less control over the actions of others than over our own action), this distribution appears to be motivated by the semantics of the gerundive complement, i.e. by its tendency to portray the complement process as being somewhat 'detached' from the influence of the main clause agent(s).

It might follow from the above that the theory of reduced-increased controllability and/or dispersal of control, as advocated in this paper, is applicable only to agentive events, which presuppose the notion of control. The question that arises at this point is how this theory relates to non-agentive events, which are coded by the infinitival complement rather the gerundive one, although they are not controllable in the strict sense of the word (cf. examples 20, 21). Yet, this distribution is not contradictory if one considers that these non-volitional process (e.g. one's hair going grey or getting old), portray the subject referent as an experiencer of the process described in the complement clause, or a thing that is directly affected by the complement process. Hence, also here the subject referent is construed as directly involved in the complement process, although it is 'experiential' rather than causative involvement. This, in turn, is in line with the theory proposed in this paper in that the degree of one's involvement in a process (in the sense of being more or less affected by it) and the degree of one's causal influence on an action appear to represent two sides of the same coin.

Finally, it must be reiterated that most of the semantic contrasts discussed in this paper are gradient rather than absolute, which is consistent with the generally accepted view that lin-

guistic choices are probabilistic in nature rather than categorical. Nevertheless the fact remains that certain construction types are strongly and significantly preferred in certain types of contexts. The distributional preferences (or dispreferences), as discussed in this paper, provide evidence for another widely accepted view, namely that full synonymy is a rare occurrence and a change in form always spells a difference in meaning, as famously argued by Bolinger (1968).

7. Concluding remarks

The corpus-based, bottom-up approach employed in this study has allowed to determine the linguistic factors that significantly affect the choice of the infinitival or gerundive complement in cases where both types are accepted by a given matrix predicate. It should be clear that one would not be able to detect all these significant patterns of use with purely introspective or intuition-based methods. Thus, the role of corpus-based methods in analysing closely related or nearly-synonymous constructions is not to be overestimated. However, it must be emphasized at this point that the identification of significant patterns of use has not been the ultimate goal of this research. Distributional patterns have been taken here to represent or to be indicative of semantic patterns, which is a particular case of the distributional hypothesis. The qualitative interpretation of the frequency data has confirmed this line of thinking, which is to say that some general principles governing and motivating the infinitive-gerund alternations have emerged from the data and have been generalized as a *reduced-increased involvement theory*. However, given the limited scope of the present study, this theory should be viewed as a working theory rather than a final or fully confirmed one. There are still many questions that need to be addressed in future research. First of all, it remains to be seen how this account relates to constructions governed by other verbs that participate in infinitive-gerund alternations – a task that could not be undertaken in this paper due to spa-

ce limitations. Hence, possibly other linguistic/contextual factors will be discovered that significantly affect the choices between the two constructional variants under analysis. The next worthwhile step seems to be the application of multivariate statistics in order to determine the predictive power of particular variables being discussed. Also, a comparison of the findings presented in this study with data from other languages might provide a worthwhile insight into the cross-linguistic validity of the theory proposed in this study.

References

- Bolinger, Dwight (1968). "Entailment and the meaning of structures". *Glossa* 2: 119–127.
- Bresnan, Joan (1979). *Theory of Complementation in English Syntax*. New York: Garland.
- Deshors, Sandra C. (2015). "A constructionist approach to gerundial and infinitival verb complementation patterns in native and Hong Kong English varieties". *English Text Construction* 8/2: 207–235.
- Deshors, Sandra C., Stefan Th. Gries (2016). "Profiling verb complementation constructions across New Englishes: A two-step random forest analysis of *ing* vs. *to* complements". *International Journal of Corpus Linguistics* 21/2: 192–218.
- Dirven, René (1989). "A Cognitive Perspective on Complementation". In: Dany Jaspers, Wim Klooster, Yvan Putseys, Pieter Seuren (eds.). *Sentential Complementation and the Lexicon*. Dordrecht: Foris Publications, 113–139.
- Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele (2006). *Constructions at Work*. Oxford: Oxford University Press.
- Givón, Talmy (1980). "The binding hierarchy and the typology of complements". *Studies in Language* 4: 333–377.
- Givón, Talmy (2001). *Syntax: An Introduction*, Vol. 2. Amsterdam: John Benjamins.

- Gries, Stefan Th., Anatol Stefanowitsch (2004). "Extending colostruc-tional analysis: A corpus-based perspectives on 'alternations'". *International Journal of Corpus Linguistics* 9/1: 97–129.
- Gries, Stefan Th. (2017). "Syntactic alternation research. Taking stock and some suggestions for the future". *Belgian Journal of Linguistics* 31: 8–29.
- Harris, Zellig S. (1954). "Distributional Structure". *WORD* 10/2-3: 146-162.
- Kaleta, Agnieszka (2012). "The English gerund vs. the to-infinitive: the case of aspectual constructions". *Selected Papers from UK-CLA Meetings*, Volume 1: 323–34.
- Kaleta, Agnieszka (2014). *English Sentential Complementation: A Usage-based Approach*. Piotrków Tryb.: Naukowe Wydawnictwo Piotrkowskie.
- Krzeszowski, Tomasz P. (1990). *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin – New York: Mouton De Gruyter.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar* Vol. 1. Stanford: Stanford University Press.
- Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar* Vol. 2. Stanford: Stanford University Press.
- Langacker, Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Noonan, Michael (1985). "Complementation". In: Timothy Shopen (ed.). *Complex Constructions*. Cambridge: Cambridge University Press, 42-140.
- Rudanko, Juhani (2017). *Infinitives and Gerunds in Recent English: Studies of Non-Finite Complements with Data from Large Corpora*. Basingstoke: Palgrave Macmillan.
- Rudanko, Juhani (2011). *Changes in Complementation in British and American English: Corpus-based Studies on Non-finite Complements in Recent English*. Basingstoke: Palgrave Macmillan.
- Rudanko, Juhani (2021). *Infinitival vs Gerundial Complementation with Afraid, Accustomed, and Prone. Multivariate Corpus Studies*. Basingstoke: Palgrave Macmillan.
- Smith, Michael B. (2008). "The semantics of complementation in English: A cognitive semantic account of two English complement constructions". *Language Sciences* 31: 360–388.
- Smith, Michael B., Joyce Escobedo (2001). "The semantics of to-infinitival vs -ing verb complement constructions in English". In: Mary Andronies, Christopher Ball, Heidi Elston, Sylvain Neuvel (eds.).

Proceedings from the Main Session in the Chicago Linguistic Society's Thirty-seventh Meeting. Chicago: Chicago Linguistic Society, 549–64.

Taylor, John, Rene Dirven (1991). *Complementation*. Duisburg: LAUD Linguistics Agency, University of Duisburg.

Stefanowitsch, Anatol, Stefan Th. Gries (2003). “Collostructions: Investigating the interaction between words and constructions” *International Journal of Corpus Linguistics* 8/2: 209-43.

Wierzbicka, Anna (1988). *The Semantics of Grammar*. Amsterdam: John Benjamins.

Agnieszka Kaleta
ORCID iD: 0000-0002-6272-7163
Institute of Literature and Linguistics
Jan Kochanowski University
ul. Słowackiego 114/118
97-300 Piotrków Trybunalski
Poland
agnieszka.kaleta@ujk.edu.pl

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

<https://doi.org/10.26881/bp.2022.1.02>

Lexical factors in non-finite complementation of *continue*

JOANNA PODHORODECKA

*Received 10.06.2022,
received in revised form 2.12.2022,
accepted 4.12.2022.*

Abstract

The study examines the usage of two non-finite complements of the verb *continue*: the -ing form and the *to*-infinitive, arguing for the importance of low level generalizations in accounting for the complement choice. The semantic import of complement constructions may vary with specific lexeme classes of both the matrix verb and the complement verb, as well as be conditioned by more general features associated with the verbs' lexical aspect. The determinants of complement choice are characterized in terms of the aspectual construal of the complement event imposed by the two alternative constructions: conceptual proximity and distance associated respectively with the -ing form and the *to*-infinitive. The study relies on distinctive collexeme analysis: a statistical technique which compares the lexemes distinctive for the two constructions in order to describe the semantics of the construction by examining its most characteristic collocates.

Keywords

non-finite complementation, aspectual verbs, construction grammar, lexical aspect, distinctive collexeme analysis

Uwarunkowania leksykalne komplementacji angielskiego czasownika *continue*

Abstrakt

Celem artykułu jest analiza dwóch alternatywnych wzorców komplementacji angielskiego czasownika *continue*: formy gerundialnej z końcówką *-ing* oraz bezokolicznika z *to*. Autorka argumentuje, że motywację wyboru przez mówiących jednej z dwóch konstrukcji może wyjaśnić jedynie analiza na niskim poziomie abstrakcji, ponieważ ich znaczenie różni się w subtelny sposób dla różnych typów zarówno czasownika nadrzędnego jak i podrzędnego, a dodatkowo jest uwarunkowane cechami natury bardziej ogólnej, związanymi z aspektem leksykalnym. Czynniki warunkujące wybór wzorca komplementacji są opisane w kategoriach konturu aspektualnego tworzonego przez dwie alternatywne konstrukcje: bliskości i dystansu pojęciowego powiązanego odpowiednio z formą z końcówką *-ing* oraz bezokolicznikiem z *to*. Badanie opiera się na analizie charakterystycznych koleksemów (ang. *distinctive collexeme analysis*): technice statystycznej porównującej leksemę dystrynktywne dla dwóch konstrukcji, w celu opisanie semantyki tych konstrukcji za pomocą ich najbardziej typowych kolokacji.

Słowa kluczowe

komplementacja czasowników angielskich, czasowniki aspektualne, gramatyka konstrukcji, aspekt leksykalny, analiza koleksemów dystrynktywnych

1. Introduction

The aim of this study is to propose a usage-based model of the non-finite complementation of *continue*, building on the lexical preferences of the two constructions revealed by distinctive collexeme analysis: a statistical technique of collocation analysis established by corpus-driven cognitive semantics (Glynn 2014: 315). Cognitive grammar (Langacker 1991, 2008) and Construction Grammar (Goldberg 1995, 2006) are used as the framework for the analysis. The study argues for the importance of low-

level generalizations in the description of verbal complementation, demonstrating that complement constructions may have different semantic import with different lexeme classes.

Recent cognitive and functional approaches to complementation have generally ceased looking for single, unified semantic determinants of complement choice, emphasizing instead the multiplicity of essentially probabilistic factors involved, which may be lexical, semantic, discourse-pragmatic, even prosodic in nature (De Smet 2013: 31-41). This study focuses on the lexical and aspectual features involved in the choice between the two complements: the interaction between the semantics of the complement verbs and the aspectual construal of the event inherent in the complement construction with the matrix verb *continue*. Construction Grammar argues that grammatical constructions are inherently meaningful (Hilpert 2021: 10), and thus the analysis of complementation patterns should encompass the semantic content contributed by all their elements (Duffley & Fisher 2021: 72) in order to examine how the schematic meaning of the construction interacts with the more specific semantic content of the complement verb. This in turn necessitates analysis at a more specific level: rather than describing characteristics of the two complement constructions in general, the present study focuses on the regularities specific to a particular matrix verb (*continue* as an aspectual verb with a specific semantic content) as well as particular semantic classes of complement verbs. The two complement constructions are here approached as a cluster of more specific constructional schemas.

The paper is structured as follows. The current section introduces the aims and scope of the study. Section 2 gives a brief overview of the differences between the two non-finite complements of *continue*: the *to*-infinitive and the *-ing*. Section 3 is devoted to a distinctive collexeme analysis of the two constructions, discussing first the top 20 complement verbs distinctive for each of them and then proposing a more detailed description of all the statistically significant collexemes divided into semantic classes. Section 4 discusses the implications of the results

and offers some conclusions pertaining both to the examined constructions and to the adopted research methodology.

2. Non-finite complements of *continue*

Non-finite complements of aspectual verbs, such as *begin*, *continue* or *cease*, pose problems for semantically motivated accounts of verbal complementation: often there is no immediately apparent difference in meaning between the alternative constructions, and the determinants of complement choice that prove explanatory for other types of verbs are of limited use or fail altogether. For instance, the three basic semantic oppositions proposed by Hamawand (2004: 452-455), temporal reference (subsequent – non-subsequent), aspect (bounded – unbounded) and character (more nominal – more verbal), cannot account for the difference between the non-finite complements of *continue*. Both patterns code the same temporal reference: the matrix event and the complement event are co-extensive. Attempts to explain the difference in terms of interruption and subsequent resuming of the activity find no support in data (Egan 2008: 192). The nominal-verbal distinction is in fact misleading when applied to the complements of *continue*, as it is the -ing complement, typically described as more nominal, that denotes actions which are more agentive and specific rather than generic (Egan 2008: 185-186). Out of the three semantic determinants of complement selection, the only one that may be applicable to *continue* constructions is aspect.

The bounded-unbounded distinction related to aspect can be described in terms of specific features of construal: perspective and viewing frame, i.e. the part of the event that is in focus. The -ing form has an imperfectivizing effect and takes internal perspective on the verbal process, so that its boundaries are excluded from the immediate perceptual scope. Since the conceptualizer 'zooms into' the event, it is viewed as a series of homogeneous states (Langacker 2008: 120-121). The *to*-infinitive signals boundedness, so the event is construed holistically, as a single unit (Hamawand 2004: 453). This distinction, however,

is blurred with *continue*, where the semantic content of the matrix verb itself imposes internal perspective on both non-finite constructions. Depending on the telicity of the complement verb, the event is consequently seen as either continuative (*continue reading this story*)¹ or iterative, i.e. denoting a series of repetitions (*continue reading novels*). The difference in perspective inherent in the aspectual distinction may in this case serve more general, discourse-pragmatic functions.

Duffley (2006: 107-115) contrasts the two complements of *continue* in terms of their syntactic function. He describes the *to*-infinitive complements as goal-circumstantial, emphasizing "the notion of movement towards the total realization of an event" (2006: 111). Rather than on the complement process itself, the construction focuses on its potential completeness and goal. By contrast, the *-ing* complement has the function of direct object, "that which is continued" (2006: 108), and it places the focus on the complement process.

The theoretical tools that this study will mostly rely on are Verspoor's (1999, 2000) analysis of non-finite complements in terms of conceptual proximity or distance and Egan's (2008) 'targeted alternative' approach to the meaning of the *to*-infinitive. Verspoor (2000: 202), following Givon (1993), argues that English complement constructions are motivated by iconic principles, in that the degree of syntactic integration of the main clause and the subordinate clause reflects the conceptual integration of the two events. Consequently, various constructions can be described in terms of conceptual proximity or distance. A close vantage point connected with the internal perspective of the *-ing* complement presents the event as experienced directly and therefore subjectively. Thus, the *-ing* form codes "a very subjectively construed atemporal imperfective process that does not include the initial and the final states" (Verspoor 1997: 449).

By contrast, the *to*-infinitive has a distancing function and portrays the complement event as experienced indirectly, where the distance may be either temporal or epistemic. Depending on

¹ All examples from COCA corpus (Davies 2008-), emphasis added.

the type of matrix verb, it may introduce the notion of futurity (with manipulative verbs, e.g. *want to*, *promise to*) or a judgment going beyond the immediate experience (e.g. *find/believe somebody to*) (Verspoor 2000: 208). Because the *to*-infinitive inherits the path-goal sense of the preposition *to*, which implies both distance and direction (Egan 2008: 95), it is often seen as goal-oriented.

Egan's (2008) description of the meaning of the infinitive in terms of 'targeted alternative' is an extension of this approach. A *to*-infinitive construction denotes "a situation, viewed as a whole, profiled as the more/most likely of two or more alternatives in some specified domain" (2008: 99). For instance, with matrix verbs involving judgment (e.g. *consider/believe somebody to*) the situation is profiled as likely to be true, while with verbs of intention (e.g. *decide/plan to*) it is construed as likely to happen.

Egan's extensive corpus-based study of non-finite complements frequently emphasizes the importance of context and relies on contextual features to characterize the constructions: *continue*-ing typically combines with agentive subjects and describes specific events (single or repeated on one particular occasion), while *continue to*-infinitive favours non-agentive subjects and denotes generic situations, likely to occur more or less regularly on different occasions (Egan 2008: 185-187). Even though the analysis that follows focuses on token frequency of particular complement verbs rather than close analysis of their individual uses in context, features such as agentivity and specificity will prove to be important factors in the interpretation of the results.

3. Distinctive collexeme analysis

3.1. The method

Distinctive collexeme analysis belongs to the family of quantitative corpus-linguistic methods jointly described as collocation analysis (Stefanowitsch and Gries 2003; Gries and Stefa-

nowitsch 2004). It is aimed at semantic description of grammatical constructions by characterizing the meaning of their collexemes: lexical items that are most strongly associated with them. In order to gauge the strength of this correlation, referred to as collocation strength, the analysis uses relative frequencies to identify lexical items which occur in a construction more often than expected (Hilpert 2014: 391-392). Distinctive collexeme analysis is specifically geared for contrasting two or more similar constructions. By comparing their collocational preferences, the method reveals subtle differences between superficially synonymous structures. The input data consist of the overall frequencies of both constructions in the corpus and the token frequencies in both constructions of each of the analysed collexemes. Thus, the distinctive collexeme analysis of the complements of *continue* requires the overall frequency of *continue to-infinitive* and *continue -ing* in the corpus, as well as the frequencies of particular complement verbs in both constructions. The results will show which of the complement verbs occur in one structure significantly more often than in the other.

The data used in the present analysis have been obtained from the COCA corpus (Davies 2008-), which contains over 500 million words and is continually updated. Due to the large size of the corpus, the sample was limited to complement verbs whose frequency of usage exceeds 0.1 % of at least one of the constructions (8 tokens for *continue -ing* and 62 tokens for *continue to-inf*). The analysis was performed with Coll.analysis 3.5 script (Gries 2014), kindly made available by the author. The script uses the Fisher exact test, producing as the final output not the resulting p-value, but its negative logarithm to the base of ten. This procedure greatly simplifies the interpretation: it avoids very low p-values, frequently expressed in powers of ten, instead providing more intuitively readable results, where a larger number corresponds to a stronger association between the lexical item and the construction.

The analysis proceeds as follows. First, the top 20 collexemes revealed as distinctive for each construction are compared with Kaleta's (2014) findings based on the BNC corpus. Then a more

detailed interpretation procedure is proposed, following Perek (2014: 115-141); namely, all statistically significant collexemes are divided into semantic classes and generalizations are made over each of such ‘verb-class-specific’ constructions.

3.2. Top 20 distinctive collexemes

Distinctive collexeme analysis of the two non-finite complements of *continue* has already been conducted by Kaleta (2014: 122-125) on the basis of the British National Corpus. The study revealed 18 significant collexemes distinctive for *continue to*-infinitive and 65 for *continue -ing*. The main complements of the infinitival construction are verbs of state (*be, have, exist*), verbs describing non-agentive processes of change and development (*rise, grow, increase*), cognition and perception verbs (*believe, stare*) and abstract, generic agentive processes (*press, support*). The *continue -ing* construction attracts verbs denoting specific, agentive actions (*work, walk, read*) and recurrent processes (*trade, collect, finance*).

The aim of the present analysis is to validate Kaleta's findings by investigating a larger corpus and hopefully to demonstrate the benefits of low level generalizations in the interpretation of the results.

The analysis of the COCA corpus data revealed 92 distinctive collexemes of *continue to*-infinitive and 87 of *continue -ing*. The top 20 complements of both constructions are listed below, together with their collocation strength. The lexemes that occur also in Kaleta's (2014: 122-125) top 20 are marked in bold.

Table 1

Top 20 significant collexemes of
continue to-infinitive and *continue -ing*.

Continue to-infinitive	Collocation strength	Continue -ing	Collocation strength
Be	Infinite	Walk	95.5381
Have	27.4413	Talk	89.7903
Grow	22.7274	Cook	84.4527

Increase	13.0338	Work	58.325
Exist	12.7109	Read	55.343
Decline	11.8842	Bake	37.5642
Insist	11.3002	Beat	27.5865
Evolve	11.0527	Add	16.5274
Maintain	10.2658	Whisk	12.9695
Believe	8.3925	Eat	12.346
Struggle	8.3212	Roast	12.335
Improve	8.2822	Live	12.1366
Dominate	7.6177	Use	11.7114
Hold	7.299	Speak	10.7778
Keep	7.2074	Fight	10.3168
Mount	6.9105	Drive	10.0814
Rely	6.4111	Sing	9.9782
Thrive	6.2684	Write	9.9148
Plague	6.1757	Officiate	9.3309
Haunt	5.8106	Fund	9.1211

The collocation strength determined for a particular verb reflects the p-value resulting from the Fisher exact test. The lexemes whose collocation strength exceeds 1.3 are statistically significant at the p-level of 0.05, while a value above 3 indicates a highly significant attraction at the p-level of 0.001. As can be seen in the table above, all top 20 lexemes are well above this threshold. The collexemes of *continue* -ing are generally more distinctive, which may be caused by the lower overall frequency of the construction.

The results show that there is a considerable degree of correspondence between the distinctive collexemes in both corpora: 17 verbs occur within the top 20 on both lists, while 16 more lexemes appear within the top 20 in one ranking, but further down the list in the other. The correspondence is more pronounced in the case of the infinitival construction, where the top five COCA collexemes appear in BNC top 20, and the two most significant ones, *be* and *have*, match in both lists. Curiously, the verb *stare*, which in the BNC appeared among the top 20 collexemes of *continue to*-infinitive, turned out to be distin-

ctive for *continue* -ing in COCA, with a collostruction strength of 2.5595, which ranks it approximately in the middle of the list.

Two of the semantic classes postulated by Kaleta (2014: 122-125) can also be observed in COCA data. State verbs (*be*, *have* and *exist*) remain the most distinctive for *continue to*-infinitive, closely followed by verbs of non-agentive change and development (*grow*, *increase*, *decline*, *evolve*, *improve*, *mount* and *thrive*). Additionally, a group of verbs can be distinguished whose common semantic trait is the notion of continued influence, either agentive or abstract (*maintain*, *dominate*, *hold*, *keep*, *rely*, *plague* and *haunt*). What the remaining 3 verbs (*insist*, *believe* and *struggle*) have in common is that they typically require human subjects and indicate effort and conviction, so they can be construed as force-dynamic. Generally, it can be concluded that the distinctive collexemes of *continue to*-infinitive are predominantly durative and atelic since they are either states or abstract activities.

The verbs distinctive for *continue* -ing indeed describe specific agentive events: motion (*walk* and *drive*), communication (*speak* and *talk*) and other durative processes, typically intransitive (*work*, *live*, *fight* and *officiate*). A characteristic group of collexemes refers to the processes of food preparation (*cook*, *bake*, *beat*, *add*, *whisk* and *roast*). Finally, there is a group of usually transitive verbs whose lexical aspect depends to a great extent on the type of the object (*read*, *eat*, *use*, *sing*, *write* and *fund*). With a singular, specific object they denote a single accomplishment – a durative, telic action (e.g. *continue writing the letter*) while with a plural object they gain iterative meaning and describe a series of accomplishments (*continue writing letters*). The common features of the collexemes of *continue* -ing seem to be dynamicity and durativity – they code either activities (dynamic, durative and atelic) or accomplishments (dynamic, durative and telic). The durativity of the *continue* constructions results directly from the semantics of the matrix verb, but the *to*-infinitive and the -ing complements can be contrasted in terms of their respectively atelic and dynamic collexemes. The problem is that these two features are not mutually exclusive: activities are at

the same time dynamic, durative and atelic, and thus constitute a potential area of overlap between the two constructions.

3.3. Verb-class-specific analysis

As can be seen above, there are certain generalizations that can be made over the groups of verbs distinctive for each construction. Casting the net wider and analysing a larger group of collexemes will hopefully reveal more clearly visible patterns related to specific semantic classes. “Many constructions are distributionally biased towards one verb whose meaning is very similar (if not identical) to that of the construction” (Perek 2014: 89) and such regularities are the most clearly visible at the lower levels of generalization. As the following analysis will reveal, in most of the verb classes discussed below, there is one dominant verb, usually twice as distinctive for the construction as the next item in the ranking, while among the top 20 significant collexemes the differences in collocation strength are much less pronounced. Such dominant verbs have been marked in bold in the tables that follow, and they are assumed to reflect the meaning of the constructional sub-schema the most closely.

The division of the collexemes proposed below is based on the verbs' semantic domain, event structure (Croft 2012) and lexical aspect (Vendler 1957; Croft 2012). The semantic criterion is treated as primary where a specific domain can be distinguished. With the remaining verbs, generalizations are sought at a higher level, on the basis of features such as durativity, transitivity or aspect.

3.3.1. Distinctive collexemes of *continue to-infinitive*

The first group of collexemes of the infinitival construction contains state verbs, which describe continued existence of a situation or continued possession of a particular quality.

Table 2Distinctive collexemes of *continue to*-infinitive: verbs of state

Be	Infinitive	Stand	3.6108
Have	27.4413	Stay	3.4202
Exist	12.7109	Represent	1.6876
Remain	5.5946	Occupy	1.3948
Show	4.7771		

The most distinctive of those collexemes is the verb *be*, whose strength of association with the construction is infinite. In the majority of its uses, the verb is followed by a subject predicative, but approximately 36 % of its tokens are examples of the passive, which in itself has a stativizing effect (Langacker 1991: 265). There are altogether 9 significant collexemes of this micro-construction (*continue to*-infinitive passive). They either coincide with the collexemes of the *continue to*-infinitive (*plague, haunt, dominate*) or resemble them semantically (*dog, debate, marginalize*), while the remaining verbs (*amaze, fascinate, frustrate*) form adjectival passives. At the same time, the construction strongly repels typical dynamic verbs (e.g. *make, do, find, give*), which proves that it inherits more features from *continue to*-infinitive than from the passive construction in general.

The verb *have* is a light verb, whose meaning greatly depends on its object. The most frequent combinations refer to experience, usually negative (*have problem/trouble/difficulty*) or attitude, usually positive (*have confidence/faith/hope*). The verb *show* is predominantly complemented either by that-clauses or abstract nouns (*show progress/growth/improvement*), to express a judgement or a conclusion. *Stand* denotes posture or attitude (*stand by/together/defiant*), while the tokens of *represent* divide almost equally between the stative meaning of 'constitute' (*represent an obstacle/a hazard*) and the dynamic of 'stand for in court'. The considerations above indicate that the classification of the verbs is necessarily an approximation, as few lexemes show entirely uniform usage. Often the interpretation is conditioned by features such as subject animacy or transitivity type.

The next group of collexemes contains verbs profiling change and development.

Table 3

Distinctive collexemes of *continue to*-infinitive:
verbs of change and development

Grow	22.7274	Fall	3.8516
Increase	13.0338	Become	3.7991
Decline	11.8842	Expand	3.2047
Evolve	11.0527	Emerge	2.8808
Improve	8.2822	Worsen	2.4564
Mount	6.9105	Unfold	2.3857
Thrive	6.2684	Rage	2.3715
Rise	5.8026	Shrink	2.2299
Flourish	4.427	Appear	2.0896
Change	4.3839	Escalate	1.8717
Occur	4.2743	Happen	1.5829
Drop	4.1951	Lose	1.304

The verbs describe change in quantity (*increase, decline, rise, drop, expand, shrink*) or quality (*improve, worsen, change, become*), occurrence (*occur, emerge, appear, happen*) or continuation of a process (*thrive, flourish, rage, unfold*). They are typically non-agentive, combining with abstract or inanimate subjects. Most of them are evaluative, specifying whether the process in question is a positive or a negative one. All these features are exemplified by the most distinctive verb *grow*. In terms of their lexical aspect, the collexemes are durative and predominantly atelic; however, some of them can be construed as telic and consequently iterative (*new cases continue to occur*).

Another large group comprises verbs of extended causation or influence.

Table 4

Distinctive collexemes of *continue to*-infinitive:
verbs of causation and influence

Maintain	10.2658	Shape	3.0088
Dominate	7.6177	Exert	2.6443
Rely	6.4111	Pose	2.363
Plague	6.1757	Generate	2.3059
Influence	5.8106	Draw	2.2348
Haunt	5.8106	Gain	2.034
Affect	5.1721	Define	2.0329
Support	4.6301	Benefit	1.9496
Attract	4.5063	Monitor	1.8693
Inspire	3.4269	Control	1.8669

They are mostly non-agentive, describing cause-effect relationships between abstract entities. Those that can occur with a human subject profile extended, relatively non-specific processes of causation and control (*maintain, support, monitor, control*). They are overwhelmingly durative and atelic.

As opposed to the previous three groups of verbs, which typically co-occur with inanimate, often abstract subjects, the following groups comprise verbs that normally require a human subject. 16 of them describe communication and interaction (see Table 5).

Table 5

Distinctive collexemes of *continue to*-infinitive:
verbs of communication and interaction

Insist	11.3002	Claim	2.3715
Struggle	8.3212	Press	2.151
Face	5.1345	Allow	2.0891
Ignore	3.7397	Express	2.0696
Deny	3.27	Emphasize	2.0164
Resist	3.5055	Refuse	1.9533
Oppose	2.9283	Challenge	1.6133
Elude	2.6915	Threaten	1.4668

The common feature of the verbs is that they are force-dynamic: they involve the notion of effort (*struggle, face, press*), counterforce (*resist, oppose, refuse*) or the lack of it (*elude, allow*), conflict (*challenge, threaten*) or a difference of opinion (*deny, insist, claim*). Many of them are telic, so in combination with *continue* they produce iterative reading: that of repeated complete acts of communication (*continue to insist, continue to deny*).

Cognition verbs are grouped in the next category.

Table 6

Distinctive collexemes of *continue*
to-infinitive: cognition verbs

Believe	8.3925	Examine	2.0696
Suffer	4.731	View	1.9912
Experience	4.1542	See	1.7387
Feel	3.7994	Hear	2.1634
Focus	3.1864	Enjoy	1.3523

They include verbs of experience (*suffer, enjoy, experience*), opinion (*believe, view*), mental process (*focus, examine*) and perception (*feel, see, hear*). Their basic function is to introduce propositions expressing epistemic or evaluative judgements or to describe complete acts of perception. The wide aspectual potential characteristic of cognition verbs is here limited by the matrix verb, which imposes a durative construal: the verbs may denote states (*continue to believe in yourself*), activities (*continue to enjoy their holiday*) or iterative achievements (*continue to hear gunshots*). They are mostly, but not exclusively, non-agentive. The lexemes in the next group in their basic senses denote possession and transfer.

Table 7

Distinctive collexemes of *continue* to-infinitive:
verbs of possession and transfer

Hold	7.299	Get	3.362
Keep	7.2074	Seek	1.5103
Find	4.2138	Bring	1.4854

This group is in fact rather diverse, overlapping with other categories, since the verbs are often used in their extended senses and denote obtaining or retaining possession of a more abstract entity. For instance, the verb *hold* frequently occurs in combinations describing extended influence (*hold sway/allure/appeal/authority*) while *keep* usually forms the resultative construction (*continue to keep crime down/people awake*), so they resemble verbs of causation and influence. *Get* approximates verbs of change and development, as it often combines with a subject predicative (*get sick/better/worse*) or forms the *get*-passive (*get vaccinated/ripped off*). *Find* is mostly used as a mental verb with the meaning of 'realize' or 'discover'. So, although the verbs are characterized by particularly varied usage, they still conform to the general pattern visible in other categories.

The last group of collexemes includes verbs of motion:

Table 8

Distinctive collexemes of *continue*
to-infinitive: verbs of motion

Come	4.0808	Follow	2.1439
Meet	2.8402	Flow	2.034
Lag	2.6443	Lead	1.6873
Leave	2.363		

Some of the verbs are telic (*come, meet, leave*) while others describe imperfective motion (*follow, flow, lead*). The characteristic trait is, however, that the motion is relational: the verbs are either deictic (*come* – the most distinctive verb in this category) or they denote directed motion relative to someone or something (*lag, follow, lead*).

Table 9 below summarizes the collexeme types of the *continue* to-infinitive construction. They have been rank-ordered according to the combined collocation strength of all their lexemes, which is taken to reflect the overall distinctiveness of each category.

Table 9
Distinctive collexemes of *continue*
to-infinitive: ranking of types

Type of collexemes	Dominant verb	Number of verbs	Average collocation strength	Combined collocation strength
State	<i>be, have</i>	9	Infinitive (7.5796625)	Infinitive (60.6373)
Change and development	<i>grow</i>	24	5.552916667	133.27
Causation and influence	--	20	4.10682	82.1364
Communication and interaction	<i>insist</i>	16	3.53886875	56.6219
Cognition	<i>believe</i>	10	3.35787	33.5787
Possession and transfer	--	6	4.17965	25.0779
Motion	<i>come</i>	7	2.541928571	17.7935

The most distinctive collexemes of *continue to*-infinitive denote states and non-agentive processes. Verbs that combine with human subjects are positioned further down the list; moreover, they are not typically agentive, but relate to concepts such as cognition, possession and motion. More typical realizations of the transitive schema involving direct physical causation are conspicuously absent from the list.

The evidence above suggests that the *continue to*-infinitive construction is motivated by the path schema, realized differently in different domains. With states, non-agentive processes and mental verbs the *to*-infinitive introduces the notion of epistemic distance. The collexemes of these categories refer to events which require assessment or interpretation, either because they involve epistemic and value judgements (*believe, improve, thrive*) or states and durative impersonal processes (*decline, influence*), which extend in time beyond the act of perception and thus cannot be directly observed and require a certain

amount of inference (Noël 2003: 18). With more agentive collexeme types, which are, however, less distinctive for the construction, the infinitive brings in the aspect of goal-orientedness and direction. Verbs of motion refer to deictic or relational movement (*come, follow*) while with verbs of interaction and communication the directionality inherent in the to-infinitive is realized as force-dynamics, since the collexemes denote forceful and directed communicative actions (*deny, insist*). The complement event is usually complete, viewed as a whole, and rather than on the process itself, the emphasis lies on its result, interpretation or evaluation.

3.3.2. Distinctive collexemes of *continue -ing*

Two semantic categories are represented among the collexemes of both constructions: motion and communication verbs. There are, however, considerable differences between the types of lexemes that they contain. Whereas the motion verbs attracted by *continue to-infinitive* express the concept of deictic or relational movement, the collexemes of the *-ing* construction focus instead on the manner of motion.

Table 10

Distinctive collexemes of *continue -ing*: verbs of motion

Walk	95.5381	Ride	3.266
Drive	10.0814	Pace	2.8258
Run	7.9468	Chase	2.1673
Move	6.3204	Rock	2.0043
Travel	4.0555	Head	1.47
Swim	3.5657	Dance	1.3913

The verbs specify the speed (*walk, run, pace*), mode (*drive, ride, swim*) or general character (*travel, chase*) of motion. They profile the path of movement rather than its goal or starting point, which is consistent with the unbounded construal produced by the *-ing* form. They are mostly non-relational, with the possible exception of *chase* and *head*, and as opposed to the collexemes

of *continue to*-infinitive, they are uniformly durative and predominantly atelic.

Verbs of communication and interaction distinctive for *continue* -ing are listed in Table 11.

Table 11

Distinctive collexemes of *continue* -ing:
verbs of communication and interaction

Talk	89.7903	Look	2.7224
Speak	10.7778	Discuss	2.7078
Sing	9.9782	Stare	2.5595
Yell	3.9805	Scream	1.8914
Chat	2.7754	Tell	1.4715
Laugh	2.7702	Describe	1.4071

Similarly to motion verbs, the communication and interaction verbs distinctive for *continue* -ing profile imperfective actions in progress, without their boundaries. In other words, the verbs focus on the process of communication rather than its content or result: the most characteristic verbs are atelic (*talk, speak, yell*) whereas more telic verbs, *tell* and *describe*, are only marginally distinctive, positioned at the bottom of the list. The verbs *look* and *stare* included in the category also follow this pattern: they are atelic and agentive, and as opposed to central perception verbs distinctive for the *to*-infinitive construction (*see, hear, feel*), they profile the process of perception rather than its content. While the collexemes of *continue to*-infinitive describe forceful interaction between the participants of a communicative act together with its interpretation (*insist, deny, refuse*), the verbs characteristic for the -ing construction closely focus on one participant only, describing in detail the type of sounds produced (*speak, sing, laugh, yell*), but not their interpretation or the intended effect on the listener. This construal results from the conceptual proximity inherent in the -ing construction: the event is directly perceived and instantly categorized at a very basic level. As opposed to the collexemes of the infinitival cons-

truction, there is no conceptual distance necessary for a more extensive interpretation.

A characteristic group of collexemes contains verbs describing food preparation.

Table 12

Distinctive collexemes of *continue* -ing: culinary verbs

Cook	84.4527	Roast	12.335
Bake	37.5642	Sauté	7.5355
Beat	27.5865	Stir	5.5376
Add	16.5274	Simmer	4.9341
Whisk	12.9695	Process	2.9138

The group is indicative of a very specific usage context for *continue* -ing: that of recipes and instructions in general. Even verbs which are not semantically limited to food preparation (*beat*, *add*, *process*) still occur in their culinary sense in the majority of tokens of the construction (respectively 82 %, 69 % and 62 %) and their remaining uses are almost exclusively in other types of instructions. The context also has characteristic syntactic features: the matrix verb is used in the imperative and there are frequent occurrences of the null-instantiation construction – the object is not explicitly mentioned as it can easily be recovered from the context (*add carrots and continue stirring* \emptyset , *continue whisking* \emptyset *until smooth*). Thus both participants, the agent and the patient, are backgrounded, while the sole focus is on the process itself. Such a combination of a specific discourse function with a particular set of syntactic features is responsible for the high distinctiveness of the verbs.

The remaining collexemes do not show such clearly distinguishable semantic categories, therefore the further divisions are based on the verbs' lexical aspect and event structure. The first group includes iterative, i.e. serial actions.

Table 13Distinctive collexemes of *continue* -ing: iterative actions

Fund	9.1211	Count	2.0574
Pay	6.2677	Pick	2.0287
Plant	5.1583	Spend	1.7755
Take	4.6504	Throw	1.7528
Send	3.7333	Publish	1.7472
Fire	3.0969	Hammer	1.6783
Raise	2.6537	Collect	1.6179
Subsidize	2.5349	Lay	1.5358
Buy	2.2173	Deliver	1.43
Alternate	2.1441		

As complements of *continue*, the verbs denote series of actions: *continue to fire* implies repeated shots, *continue to buy* – repeated purchases and *continue to alternate* – repeated changes. Many of the verbs involve the notion of transfer (*take, send, throw, deliver*) or financial transactions (*pay, buy, spend, collect*). Others describe short repetitive physical actions (*fire, pick, hammer*).

A larger group of verbs denote durative actions.

Table 14Distinctive collexemes of *continue* -ing: durative actions

Work	58.325	Play	5.2963
Live	12.1366	Learn	4.036
Use	11.7114	Study	2.8004
Fight	10.3168	Research	2.4041
Officiate	9.3309	Abuse	2.33
Search	9.0592	Coach	1.9586
Operate	7.6914	Test	1.8263
Teach	7.4716	Campaign	1.6783
Try	6.2802	Practice	1.534
Excavate	5.8114	Compete	1.3487

The verbs include some common intransitives (*live, work, play*), lexemes denoting long term processes of transmitting or acquiring skills and knowledge (*learn, study, teach, coach, practice, re-search*) and verbs of extended forceful interaction implying conflict or competition (*fight, officiate, campaign, compete*). As activities, they are dynamic, atelic and durative, so their common features are agentivity and the continuative construal of the event.

The final group of collexemes comprises accomplishments.

Table 15

Distinctive collexemes of *continue* -ing: accomplishments

Read	55.343	Cut	3.8419
Eat	12.346	Make	2.9858
Write	9.9148	Build	2.9651
Paint	7.2196	Do	2.8055
Film	6.9615	Pull	2.7761
Wrap	5.8012	Dig	1.9005

Accomplishments are dynamic and durative, but as opposed to activities, they are telic. The verbs above are syntactically transitive and what is more, they approximate the semantic transitive prototype: they denote specific physical actions, have agentive subjects and objects whose affectedness determines the completeness of the event. Characteristically for this category, the actions can be construed either as durative or iterative, depending on the boundedness and specificity of the object (*continue reading this story* vs. *continue reading novels*). If the verbs are used intransitively, the action can also be construed as continuative (*continue reading aloud*).

The distinctive collexemes of *continue* -ing construction are rank-ordered in Table 16.

Table 16Distinctive collexemes of *continue* -ing: ranking of types

Type of collexemes	Dominant verbs	Number of verbs	Average collocation strength	Combined collocation strength
Culinary verbs	<i>cook</i>	10	21.23563	212.3563
Durative actions	<i>work</i>	20	8.16736	163.3472
Motion	<i>walk</i>	12	11.71938333	140.6326
Communication and interaction	<i>talk</i>	12	11.06934167	132.8321
Accomplishments	<i>read</i>	12	9.57175	114.861
Iterative actions	--	19	3.010594737	57.2013

The groups of collexemes most characteristic semantically are culinary verbs and motion verbs, followed by verbs of communication and interaction. As opposed to the collexemes of the infinitive construction, the majority of them describe directly perceived physical actions, with the emphasis on their exact manner. In terms of lexical aspect, the verbs' most distinctive feature is durativity. The collexemes of *continue* -ing can be both telic and atelic, but their significance decreases as telicity increases: the two telic categories, accomplishments and iterative actions, are the lowest in the ranking. The collexemes of *continue* -ing are more distinctive in comparison with the infinitival construction and the dominant verbs in each group are much more characteristic. At the same time, the collexemes are less distinct from each other: while the lexemes characteristic for *continue to*-infinitive divide into agentive and non-agentive ones, the verbs distinctive for the -ing construction are uniformly agentive, dynamic and durative.

4. Conclusions

The results of the distinctive collexeme analysis yield themselves to the interpretation in terms of conceptual proximity and distance. *Continue* -ing codes conceptual proximity and it is used for activities that can be directly experienced and immediately categorized, while *continue to*-infinitive implies epistemic distance and denotes events which require assessment or interpretation, either because they involve value judgements (*worsen, improve*), complete speech acts (*claim, refuse*) or states and durative impersonal processes, whose categorization requires some inference (*affect, support*). The distinction between conceptual proximity and distance is also visible in the specificity and lexical complexity of the verbs. The collexemes of *continue* -ing are predominantly basic level terms, describing physical actions and associated with specific mental images. They stand in stark contrast to the collexemes of the infinitive construction, which are much more abstract and generic. Even the length of the lexemes differs: among the verbs distinctive for *continue* -ing: 77 % are one syllable words, while *continue to*-infinitive has only 45 % of one-syllable collexemes.

Collexeme analysis at the level of individual semantic classes appears to be a promising method: it reveals regularities connected not only with the interplay of a specific type of matrix verb with the two complement constructions, but also patterns related to particular semantic classes of complement verbs, whose statistically significant associations with either of the two alternative constructions provide evidence for the semantic content of the constructions themselves. As construction grammarians argue, if constructions were just meaningless formal templates, there would be no visible collocational preferences between syntactic structures and specific lexical items (Stefanowitsch and Gries 2003: 236). As the match between them is evidently not random, the divergent lexical preferences of the two *continue* constructions can be assumed to reflect contrasts in their respective aspectual construal of the complement event.

Particularly interesting regularities can be noticed in the situations where two different complement types associate with similar groups of complement verbs, namely in the case of communication verbs and motion verbs. The fine but fairly regular contrasts between the semantic content of the collexemes reveal subtle differences in the semantic import of the two constructions: the to-infinitive complement associates with verbs denoting deictic or relational movement and forceful, content-oriented speech acts, whereas the -ing complement puts emphasis on the manner of the action – motion or elocution – viewed in extreme close-up. This difference can be illustrated by juxtaposing the single most characteristic verbs in each group: *continue talking* vs *continue to insist* and *continue walking* vs *continue to come*. The notions of proximity and distance, combined with the distinctions related to the verbs' lexical aspect, such as telicity or durativity, have thus proved to be useful in describing the semantic consequences of different aspectual construal of the two complement constructions examined.

There are of course some limitations of the collostructional approach, stemming from the fact that it deals with lexeme types, not tokens (Glynn 2014: 309-310). A lexeme, especially a verbal one, is a whole set of senses and usage patterns. The meaning of a verb may vary greatly with a different aspectual construal or argument structure, which is evidenced in the analysis above by the difficulties in assigning certain verbs to semantic classes due to their varied usage. That is why making generalizations requires a great deal of caution and close reference to corpus data. Thus the results of the present analysis will need to be validated and expanded in the course of further studies with the use of multivariate statistical techniques, which will closely examine a sample of individual tokens of the two constructions.

References

- Croft, William (2012). *Verbs: Aspect and Causal Structure*. Oxford: Oxford University Press.
- Davies, Mark (2008–). *The Corpus of Contemporary American English* (COCA): 410+ million words, 1990-present. Available at <<http://www.american.corpus.org>>
- De Smet, Hendrik (2013). *Spreading Patterns: Diffusional Change in the English System of Complementation*. New York: OUP.
- Duffley, Patrick J. (2006). *The English Gerund-participle: A Comparison with the Infinitive*. New York: Peter Lang.
- Duffley, Patrick, Ryan Fisher (2021). “To-infinitive and gerund-participle clauses with the verbs *dread* and *fear*”. *Studia Linguistica* 75/1: 72–96.
- Egan, Thomas (2008). *Non-finite Complementation. A Usage-based Study of Infinitive and -ing Clauses in English*. Amsterdam – New York: Rodopi.
- Givon, Talmy (1993). *English Grammar. A Function-based Introduction*. Vol 2. Amsterdam – Philadelphia: John Benjamins.
- Glynn, Dylan (2014). “Techniques and tools. Corpus methods and statistics for semantics”. In Dylan Glynn, Justyna A. Robinson (eds.). *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam – Philadelphia: John Benjamins, 307–341.
- Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.
- Gries, Stefan Th. (2014). Coll.analysis 3.5. A script for R to compute colostruational analyses.
- Gries, Stefan Th., Anatol Stefanowitsch (2004). “Extending colostruational analysis: A corpus-based perspective on ‘alternations’”. *International Journal of Corpus Linguistics* 9/1: 97–129.
- Hamawand, Zeki (2004). “Determinants of complement clause variation in English”. *English Studies* 5: 451-464.
- Hilpert, Martin (2014). “Colostruational analysis: Measuring associations between constructions and lexical elements”. In Dylan Glynn, Justyna A. Robinson (eds.). *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam – Philadelphia: John Benjamins, 391–404.

- Hilpert, Martin (2021). *Ten lectures on Diachronic Construction Grammar*. Leiden – Boston: Brill.
- Kaleta, Agnieszka (2014). *English Sentential Complementation. A Usage-based Approach*. Piotrków Trybunalski: Naukowe Wydawnictwo Piotrkowskie.
- Langacker, Ronald (1991). *Foundations of Cognitive Grammar*. Vol. 2. *Descriptive Application*. Stanford: Stanford University Press.
- Langacker, Ronald (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Noël, Dirk (2003). “Revisiting the passive of infinitival perception verb complements”. *Studia Neophilologica* 75: 12–29.
- Perek, Florent (2014). “Rethinking constructional polysemy: The case of the English conative construction”. In Dylan Glynn, Justyna A. Robinson (eds.). *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam – Philadelphia: John Benjamins, 61–86.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>
- Stefanowitsch, Anatol, Stefan Gries (2003). “Collostructions: investigating the interaction between words and constructions”. *International Journal of Corpus Linguistics* 8/2: 209–243.
- Vendler, Zeno (1957). “Verbs and times”. *The Philosophical Review* 66/2: 143–160.
- Verspoor, Marjolijn (1997). “The story of -ing: A subjective perspective”. In Martin Pütz, Rene Dirven (eds.) *The Construal of Space in Language and Thought*. Berlin: Mouton de Gruyter, 417–454.
- Verspoor, Marjolijn (1999). “To infinitives”. In Leon de Stadler, Christoph Eyrych (eds.) *Issues in Cognitive Linguistics: 1993 Proceedings of the International Cognitive Linguistics Conference*. Berlin: Mouton de Gruyter, 505–526.
- Verspoor, Marjolijn (2000). “Iconicity in English complement constructions: Conceptual distance and cognitive processing levels”. In Kaoru Horie (ed.), *Complementation: Cognitive and Functional Perspectives*. Amsterdam: John Benjamins, 199–225.

Joanna Podhorodecka
ORCID iD: 0000-0002-3131-5617
Institute of English Studies
Pedagogical University of Cracow
ul. Karmelicka 41
31-128 Kraków
Poland
joanna.podhorodecka@up.krakow.pl

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

<https://doi.org/10.26881/bp.2022.1.03>

**You shall know a term by the company it keeps:
Collocations of the term *evidence*
in general and legal corpora**

TOMASZ MICHTA

*Received 20.02.2022,
received in revised form 29.11.2022,
accepted 30.11.2022.*

Abstract

Despite the progress made in the study of collocations, their use in specialised languages by and large continues to be underresearched. The present article attempts to go some way towards filling this gap by looking at variation in collocations of a single term (*evidence*) as extracted from a general corpus and a legal one, and by exploring the implications of such variation for the retrieval of legal collocations. In particular, the study looks at a) the overrepresentation of collocations in the legal corpus, b) the underrepresentation of collocations in the legal corpus, and c) the potential of both corpora for collocation retrieval. The findings suggest that there are striking differences between the use of collocations in each corpus and that such differences can radically affect the lists of collocations obtained from each corpus.

Keywords

collocation, corpus, general language, legal language, specialised language, terminology

**Specyfika łączliwości wyrazowej terminu:
Kolokacje terminu *evidence*
w korpusie ogólnym oraz prawniczym**

Abstrakt

Pomimo postępu, jaki dokonał się w badaniach nad kolokacjami, ich użycie w językach specjalistycznych jest w znacznej mierze nadal niewystarczająco zbadane. W niniejszym artykule podjęto próbę częściowego wypełnienia stwierdzonej luki w badaniach poprzez analizę zróżnicowania kolokacji pojedynczego terminu (*evidence*) pozyskanych z korpusu języka ogólnego oraz języka prawniczego. W szczególności w badaniu zanalizowano: a) zwiększony udział niektórych kolokacji w korpusie prawniczym, b) zmniejszony udział niektórych kolokacji w korpusie prawniczym, c) potencjał obu korpusów jako źródła do pozyskiwania kolokacji. Uzyskane wyniki wskazują na istnienie zdecydowanych różnic pomiędzy użyciem kolokacji w obu korpusach, co w sposób radykalny wpływa na listy kolokacji otrzymane na ich podstawie.

Słowa kluczowe

język ogólny, język prawa, język specjalistyczny, kolokacja, korpus, terminologia

1. Introduction

Much of today's research into collocations owes a debt to the work of Firth, who is often recognised as the father of collocation. Not only did he propose *collocation* as a linguistic term, but he also succeeded in drawing the attention of numerous scholars to the habitual company that words keep. His often-quoted sentence "You shall know a word by the company it keeps" (1968:

179) captures the essence of his approach: collocations are seen as being unique to individual words and as a key aspect of a word's meaning. Following in his footsteps, many linguists have investigated the phenomenon of collocations by adopting a variety of approaches, pursuing a range of different purposes, and producing a body of literature that is impressively vast.

The motivation behind exploring collocations is at least two-fold. First, they are of theoretical interest and can be invoked to explain the fact that certain words tend to habitually occur in the company of other words (see e.g. Firth 1957, Sinclair 1991, Hoey 2005). Seen from this perspective, collocations appear to be a perfect illustration of the non-random nature of language (Kilgarriff 2005). Consequently, they need to be studied as an important subject in and of themselves, but also one that contributes to our understanding of how language works in general. Second, and related to the first, the centrality of collocations in language and the concomitant challenges that they pose for speakers act as a driving force behind undertaking more practically-oriented investigations that aim to explore the nature of the problem and contribute to overcoming it (for an overview see Boers and Webb 2017).

The challenges related to the use of collocations are by no means limited to the context of general language. Saber et al. (2020: 106) have found that problems with collocations account for the largest proportion of errors made in scientific abstracts written by low-proficiency doctoral students. Their findings tally with opinions about the difficulty of using collocations expressed by a group of professional and non-professional translators (i.e. physicians) of medical texts in a study by Badziński (2019: 167-168) who, with the use of a questionnaire, established that it was collocations that his respondents considered to be the most challenging aspect of medical translation. Some scholars also point out that, outside general language tasks, native speaker competence may be of limited use as far as collocations are concerned (Benson 1989: 4, Frankenberg-Garcia 2018). Citing her own experience as a legal translator trainer, Giczela-Pastwa (2021: 191) opines that when “starting a course

in legal translation, students are usually not familiar with legal discourse in their native tongue". This lack of familiarity with legal discourse encompasses legal collocations.

Given the importance of collocations in specialised language and the problems related to their use, it is perhaps regrettable that there is a clear imbalance in the attention paid by scholars to collocations in general language (i.e. language for general purposes, or LGP) and those in specialised language (i.e. language for special purposes, or LSP). In contrast to the wealth of insights produced by studies into collocations in LGP, until relatively recently there had been few studies into collocations in LSP. Exceptions include: Picht 1987; Martin 1992; Meyer and Mackintosh 1994; Gledhill 2000; L'Homme 2000; Heid 2001; Michta 2007; and Ward 2007. In the particular case of legal English, research by Kjaer (1990a, 1990b), Goźdz-Roszkowski (2011) and Biel (2012, 2014) merits special attention. Commenting on the state of research into legal phraseology, Goźdz-Roszkowski and Pontranfoldo (2015: 130) note: "The legal domain and its phraseology has also received scant attention". However, although this statement was true when made, the situation has since started to improve.

Also regrettable is the fact that precious few specialised dictionaries actually contain word combinations (L'Homme and Leroyer 2009: 260), while those lexicographic resources that do deign to provide them often do so inconsistently (Montero-Martinez and Buendia-Castro 2012) and may prove wanting, especially as far as their treatment of verb + term combinations is concerned (Buendia-Castro Faber 2015). The problem of inadequate lexicographical treatment of specialised collocations is further compounded by the fact that there is only a handful of dictionaries and glossaries of specialised collocations.

2. Aims and scope

The present work is of an exploratory nature and seeks to make a contribution to the study of specialised (here: legal) collocations by looking at variation in collocations that feature a single

term between a general corpus and a legal one, and by exploring the implications of such variation for the retrieval of legal collocations. More specifically, the article sets itself a three-fold aim that consists in:

- (a) establishing and exploring the most overrepresented collocations of the noun *evidence* in judgments handed down by the UK Supreme Court as opposed to general English;
- (b) establishing and exploring the most underrepresented collocations of the noun *evidence* in judgments handed down by the UK Supreme Court as opposed to general English;
- (c) investigating the potential of two different corpora (a legal one and a general one) for the retrieval of legal collocations, which is carried out with a view to including such collocations in lexicographical resources.

Thus this study represents another addition to the line of research presented in the recently published book chapter by L'Homme and Azoulay (2020), which compared sets of collocates for 15 lexical items retrieved from a) a specialised corpus on the theme of the environment and b) a general language corpus, and found marked differences between the sets retrieved from each corpus. Simultaneously, the article fits into what Biel (2010: 4) classifies as the first trajectory of research in corpus-based studies of legal language, a trajectory which focuses on external variation and investigates how “legal language differ[s] from general language and other languages for special purposes”.

Several motivations have informed the choice of the term *evidence* for the present analysis. First, the word is reasonably frequent in both corpora, occurring in them a total of 8,224 (the legal corpus) and 20,336 (the general corpus) times. This means that even though the word can act as a legal term, it may be claimed to be part of general vocabulary. Second, its significance for the field of law is unquestionable and it was felt that priority should be given to a key legal term rather than one of marginal importance. Third, its sense structure is relatively straightforward. By way of example, the Oxford Learner's Dictionaries website only distinguishes between two senses of the

word: 1. “the facts, signs or objects that make you believe that something is true” and 2. “the information that is used in court to try to prove something” (OLD 2021). Clearly, both senses are related to each other. The second is the more legal one, although its connection to law is not indicated with a label, which in turn suggests that is not perceived as being particularly technical in general language contexts. In the legal corpus chosen for the purposes of this study, the technical meaning overwhelmingly dominates and it is captured by a definition that foregrounds its field-specific character: “[t]estimony and production of documents and things relating to the facts into which the court enquires and the methods and rules relating to the establishing of those facts before the court” (Richards and Curzon 2011: 181). Importantly, *evidence* does not acquire a range of new senses¹ in the legal corpus. The existence of a single dominant meaning thus prevented the analysis from being compounded by problems of polysemy, which legal English is notorious for (Matilla 2012: 30).

3. Data and methods

3.1. Data

For the purposes of this study, two language corpora have been selected: one general and one legal. The legal corpus comprises judgments handed down by the Supreme Court of the United Kingdom (UKSC), which acts as “the final court of appeal for all United Kingdom civil cases, and criminal cases from England, Wales and Northern Ireland” (UKSC 2021a). Originally compiled by the present author and Katarzyna Mroczyńska as part of a joint project aiming to provide a lexicographic description of collocations in the judgments of the UKSC (Mroczyńska 2020,

¹ One may of course split the meaning captured by the quoted definition into a few senses. *Black’s Law Dictionary* (Garner 2004: 595), for example, lists 5 senses in its entry for *evidence*. The differences between them, albeit important whenever a high degree of precision is required, represent various focus points rather than wildly dissimilar concepts.

Michta and Mroczyńska 2022), the corpus (hereinafter: UKSCC) is intended as a reliable foundation for corpus studies of legal English. Featuring 636 actual judgments handed down by the UKSC, totalling 9.5 million tokens, and spanning a ten-year period from the Court's inception in 2009 to 2018, it represents a complete collection of UKSC judgments for the time frame mentioned. Arguably, its focus on a single text type and the rather small number of justices (12 at any one time (UKSC 2021b)), who contribute to the linguistic variety of the corpus, might be seen as a potential drawback. Yet, UKSC judgments are undoubtedly legal in nature and seem sufficiently suitable for the purpose of the study despite any linguistic peculiarities that they show compared to other legal English texts types.

As the general language corpus, the British National Corpus (BNC) was chosen to serve as a source of linguistic data. There are a number of arguments for making this particular choice. As its very name suggests, the corpus focuses on British English, thus matching the legal one in terms of the language variety it represents. Since collocations do vary between geographical language varieties (Mair 2007), ensuring this correspondence was seen as a priority. To match the legal corpus in terms of register (written), only the written subcorpus (hereinafter: WBNC) was selected for analysis. Another important consideration supporting the choice of the WBNC was its reasonably large size. Its number of tokens stands at 100.5 million (as calculated by Sketch Engine), which was considered sufficient given the study's objectives. The WBNC can also be described as balanced as it "contains texts from a wide range of different language genres and text domains" (Baker, Hardie, and McEnery 2006: 18), including a selection of academic and non-academic texts that are concerned with law. Last but not least, the WBNC is readily available online and can, for example, be downloaded free of charge or consulted via Sketch Engine. All the arguments cited so far weigh heavily in favour of using the WBNC, yet a certain drawback must also be mentioned. No texts were included in the WBNC after the release of the BNC in 1994. In fact, about 96 % of its content is comprised of texts from the period 1984-

1993.² This slightly diminishes the value of the WBNC as a reference point for the legal corpus, as the latter draws on more up-to-date texts. It was nevertheless decided that the benefits of relying on the WBNC outweigh this drawback, and any others, and it was therefore chosen as a reference corpus.

3.2. Methods

To achieve its aims, the study adopted an understanding of collocations that involves both frequency-based and syntactic criteria. In order for a word combination to potentially qualify as a collocation, the minimum frequency threshold in either of the corpora was set at 5. This rather liberal approach resulted in a rather long initial list of collocations, but all of the collocations that were subjected to closer inspection, and are discussed later in this paper, occurred more frequently in the corpus. As a safeguard against idiosyncratic collocations, another requirement was introduced so that a word combination had to appear in at least two texts in order to be considered for inclusion. Again, the approach adopted here as to what should count as a collocation contributed to the rather liberal character of the list of collocations. The final criterion was that constituents of a collocation, i.e. *evidence* (the node) and the collocates, had to be syntactically related. Only one such relation was analysed, i.e. modifier + noun. Several scholars (Bergenholtz and Tarp 1994; Michta et al. 2009: 93; L'Homme and Azoulay 2020: 154) have noted that focusing on this relation may also yield combinations that are terms, but distinguishing between collocations on the one hand, and terms on the other, was not the main focus of the study. Hence both terms and collocations were included in the analysis of collocates of *evidence*. Whenever the terminological status of certain combinations was deemed worthy of mention, the *Longman Dictionary of Law* (Richards and Curzon 2011) was consulted to establish whether a given combination typically acts

² The percentage was calculated using the BNC available via Sketch Engine.

as a term. This method, which involved taking recourse to legal resources, was felt to be a good heuristic vis-a-vis the fact that no universally agreed criteria exist that allow a clear-cut dividing line to be drawn between terms and collocations (see also Heid 2001: 791). To retrieve collocations, Sketch Engine was used since it offered all the necessary tools to ensure that all the requirements of collocations were met. In particular, the study relied heavily on Sketch Engine's word sketch functionality, which provides a condensed description of a word's grammatical and collocational behaviour (Kilgarriff et al. 2014: 9). The results produced by Sketch Engine were not accepted blindly. When a word was suggested as a potential collocate and it turned out not to act as a modifier in the corpus, the word was removed from further analysis.

In order to establish the most overrepresented collocations of the noun *evidence* in the UKSCC as opposed to WBNC, a list of key collocations was produced using the word sketch function, with the UKSCC acting as a focus corpus and the WBNC acting as a reference corpus. These key collocations were identified by Sketch Engine, which uses the simple maths method (Kilgarriff 2009). This approach calculates the keyness score according to the formula below (Lexical Computing 2015 : 3):

$$\frac{fmp(foc) + n}{fmp(ref) + n}$$

where $fpm(foc)$ is the normalised (per million) frequency of the collocation in the focus corpus, $fpm(ref)$ is the normalised (per million) frequency of the collocation in the reference corpus, and n is the smoothing parameter ($n = 1$ is the default value).

What follows from this formula is that, in general, a value of 1 indicates that a given collocation has the same normalized frequency in both the focus and the reference corpora. If it is higher than one, in general it indicates that a given collocation has a higher normalised frequency in the focus corpus (i.e. it is overrepresented there). By contrast, a value lower than 1 indicates that a given collocation has a lower normalised frequency

in the focus corpus (i.e. it is underrepresented there). It is worth noting at this point that a consequence of adopting the simple maths formula for keyness scores is that a score of 0.5 represents the same degree of underrepresentation as does a score of 2.0 regarding overrepresentation.

In order to establish the most underrepresented collocations of the noun *evidence* in the UKSCC as opposed to the WBNC, the same procedure was used. The only difference lay in the fact that the roles of the corpora (focus and reference) were reversed in Sketch Engine and the keyness score was calculated by the author using the formula mentioned earlier.³ This modification enabled a list to be produced that also included cases where the UKSCC did not contain a single occurrence of a given collocation but the WBNC did. These cases would have been absent, had the reversal not taken place as Sketch Engine removes from a list of key collocations those candidates in the focus corpus that are not featured in it. However, collocations that are attested in the WBNC and not in the UKSCC also merit inclusion in the analysis as they constitute a prime example of variation in collocation use between corpora.

In order to investigate the potential of the WBNC and the UKSCC for the retrieval of legal collocations, two word sketches were generated: one for the UKSCC and one for the WBNC. The resulting lists of collocates were sorted according to the frequency⁴ of the collocates and then compared. Unlike in the previous two stages, certain candidate collocates suggested by Sketch Engine were removed from the analysis. Such exclusion concerned (semi-)determiners (*first, other, own, and such*) as well as the adjective *only*. All of them exhibit low informative value and are typically omitted from dictionaries, terminological

³ For the sake of consistency, when calculating the keyness score, the original roles of both corpora were left unchanged, i.e. the UKSCC acted as a focus corpus and the WBNC acted as a reference corpus.

⁴ As pointed out by several authors (e.g. Łukasik 2017: 53, Michta 2018: 50, Rzepkowska 2021: 280), frequency may not be the ideal criterion for retrieving words and collocations from corpora for pedagogically-oriented publications. It is, however, an important one and provides a good starting point for analysing collocations.

glossaries or pedagogically-oriented lists of collocations. Their exclusion seemed all the more uncontroversial as, to all intents and purposes, they are unlikely to be what a user hopes to find in such resources.

4. Analysis and results

The structure of this section follows the order of aims presented earlier. First, the results of the analysis of the most overrepresented collocations of the noun *evidence* as opposed to the WBNC are presented, followed by the study's findings as to the most underrepresented collocations of the same term. Finally, the potential of both corpora for legal collocations retrieval is evaluated.

An initial list of potentially overrepresented collocations in the UKSCC featured 105 candidate collocates. The list was refined manually following the procedure described in the methods section. For reasons of space, a complete list of candidate collocates will not be presented here.⁵ Instead, only those parts are provided that are key to the analysis.

The first observation that can be made is that the UKSCC shows a marked tendency to overrepresent certain collocates. As many as 41 received a keyness score of 2 or more. The data concerning them are included in Table 1.

Table 1

The most overrepresented collocates of *evidence* in the UKSCC as opposed to the WBNC

No	Collocate	Raw frequency in the UKSCC	Raw frequency in the WBNC	Normalised frequency in the UKSCC	Normalised frequency in the WBNC	Keyness score
1	expert	188	56	19.79	0.56	13.70
2	oral	158	72	16.63	0.72	10.28
3	fresh	123	52	12.95	0.52	9.19

⁵ The complete list is available from the author upon request.

4	relevant	78	24	8.21	0.24	7.44
5	hearsay	56	30	5.90	0.30	5.31
6	such	123	178	12.95	1.77	5.04
7	other	121	178	12.74	1.77	4.96
8	cogent	33	3	3.47	0.03	4.34
9	new	87	161	9.16	1.60	3.91
10	medical	66	116	6.95	1.15	3.69
11	live	23	1	2.42	0.01	3.39
12	credible	21	3	2.21	0.03	3.12
13	anonymous	20	0	2.11	0.00	3.11
14	reliable	27	26	2.84	0.26	3.05
15	admissible	25	20	2.63	0.20	3.03
16	criminal	21	7	2.21	0.07	3.00
17	documentary	46	100	4.84	0.99	2.93
18	statistical	26	31	2.74	0.31	2.86
19	inadmissible	19	6	2.00	0.06	2.83
20	opinion	19	6	2.00	0.06	2.83
21	prima facie	25	29	2.63	0.29	2.82
22	additional	25	30	2.63	0.30	2.80
23	unchallenged	17	2	1.79	0.02	2.74
24	sufficient	53	144	5.58	1.43	2.71
25	factual	19	14	2.00	0.14	2.63
26	objective	24	35	2.53	0.35	2.62
27	false	19	15	2.00	0.15	2.61
28	material	18	15	1.89	0.15	2.52
29	character	15	3	1.58	0.03	2.50
30	only	28	65	2.95	0.65	2.40
31	general	14	4	1.47	0.04	2.38
32	conclusive	34	93	3.58	0.92	2.38
33	DNA	13	0	1.37	0.00	2.37
34	own	17	20	1.79	0.20	2.33
35	compelling	16	21	1.68	0.21	2.22

36	post-published	11	0	1.16	0.00	2.16
37	written	18	35	1.89	0.35	2.15
38	insufficient	24	70	2.53	0.70	2.08
39	further	65	279	6.84	2.77	2.08
40	available	33	118	3.47	1.17	2.06
41	primary	12	13	1.26	0.13	2.00

In Table 1, it can be seen that the overwhelming majority of collocates which are overrepresented in the UKSCC are constituted by lexical words (also called content words). This is in line with the word sketch algorithm, which prevents grammatical words (also called function words) from being displayed as collocates. Yet some of the candidate collocates may also be classified as grammatical words. The word *such*, for instance, which is one of the most overrepresented collocates, may be more suitably described as a semi-determiner (Biber et al. 1999: 281). *Other* and *own* are also grammatical words and may be classified as determiners (Biber et al. 1999: 258, 271). Determiners and semi-determiners are unlikely to be counted as collocations in pedagogically-oriented resources and theoretically-oriented analyses (especially within the phraseological strand of collocation research, see e.g. Nesselhauf 2004: 11-18) as the fact they are used with a noun is more easily explained by referring to grammatical rules than to the phenomenon of collocations. While *such* could be dismissed as being largely irrelevant for the study of collocations, the fact it placed sixth in the list seems unlikely to be accidental. Rather, it appears to reflect a tendency observed in the UKSCC to employ *such* particularly often. This word can be used to mean “of a particular of similar type” (CD 2021), as it often does in general English as well. But it can also be used to mean “this specific person/thing”, in which case it is typical of legalese (Garner 1995: 849) and is sometimes described as an anachronism (Tiersma 2000: 91) together with *aforementioned*, *same* and *said*, when used before a noun to serve deictic purposes. A concordance search revealed that it is clearly the first use that dominates in the UKSCC and con-

tributes most to its increased frequency.⁶ A particularly important function of *such*, as used in the UKSCC, is that it enables discussions of legal matters to abstract away from aspects of particular evidence to the more general aspects of categories of evidence, thus also facilitating a link between legal norms and their application to particular cases.

As regards the lexical collocates included in the table, it seems clear that a substantial number of the resulting word combinations with *evidence* exhibit different degrees of terminological character. The following are given entry status in the *Longman Dictionary of Law* (Richards and Curzon 2011): *expert evidence*, *oral evidence*, *relevant evidence*, *hearsay evidence*, *documentary evidence*, *opinion evidence*, *prima facie evidence*, *character evidence*, *conclusive evidence*, *insufficient evidence*, and *primary evidence*. The list of terms could also be extended by adding *sufficient evidence*, which unlike *insufficient evidence* is not included in the dictionary, as well as *admissible evidence* and *inadmissible evidence*, as these are related to *admissibility of evidence*, which appears as an entry.

The presence of so many terms in the list can be explained by the nature of the two corpora used for comparison. It is only natural to expect a specialised corpus intended to represent a special language known for its rich terminology (legal English) to differ from a general corpus by containing a particularly high proportion of terms, which a general corpus by definition cannot exhibit. It may be, however, no easy task to pinpoint exactly which combinations function as terms in a particular LSP and what they actually mean since this is an area where general language competence may not suffice.

The legal terms included in Table 1 differ both with regard to their transparency and their recognisability as terms. The word

⁶ The UKSCC contains sporadic instances of the construction *such evidence as (there) is/was*, some of which are found in quotes used in the judgments, e.g. “She did, however, observe that ‘such evidence as was before the Judge’ suggested that the expenditure would not have enhanced the value of the property, albeit without identifying what evidence she had in mind.” [2017] UKSC 21. In cases like this the use of *such* comes close to the second function of *such* described in the text and can be seen as replacing “the”.

oral, with the second highest keyness score, forms a term with the word *evidence*. The *Longman Dictionary of Law* (Richards and Curzon 2011: 184) states that this is evidence: “given in court by word of mouth. It may be testimony (i.e., what the witness perceived through his senses) or hearsay”. The resulting meaning, as captured by the definition, is not so far removed from the meanings of its constituents used in general language so as to make the term incomprehensible. It could therefore be assumed to be relatively transparent. At the same time, it should be noted that while the meaning of a term might appear simple after it has been explained, it is often more challenging to guess the exact meaning by relying only on the knowledge of its constituents. A number of possible meanings could be suggested for the combination of *oral* and *evidence* based on what they mean in general language (or in special language for that matter). Such predictions may prove of limited use in specialised fields such as law, where there is a marked tendency and a considerable need to ensure a high level of precision by carefully delineating a concept from similar ones (Tiersma 2008: 21). In the specific case of *oral evidence*, one might be tempted to speculate as to what it means in law and since this is a term that is relatively transparent, the results of such speculation could even be close to the definition quoted earlier. However, it seems implausible that a person uninitiated with legal English terminology would be able to suggest that the term “includes evidence which, by reason of any disability, disorder or other impairment, a person called as a witness gives in writing or by signs or by way of any device” (Criminal Justice Act 2003, s. 140). In this regard, the example of *oral evidence* shows that terms might be deceptively simple.

Full transparency of meaning cannot be assumed for any of the terms in Table 2 (or any other terms), as it is in the nature of combinations of words acting together as terms to show at least a certain degree of opacity – something which, in fact, is part of what makes them terms. Certain terms, however, can at least more easily be recognised as such. A good example is provided by *prima facie evidence*, which ranks 21 in the table and

may be immediately suspected of being a term due to its Latin constituent, even if its meaning may be impossible to guess by a person unfamiliar with Latin. Rather confusingly, the term can be employed in two different senses. While it can be used to refer to “evidence adduced by a party [which] is so weighty that it could reasonably justify a finding in his favour” (Keane and McKeown 2012: 30), it can also mean “evidence adduced by a party which is, in the absence of contradictory evidence, so weighty that it does justify a finding in his favour” (Keane and McKeown 2012: 30). Even for a person familiar with the Latin phrase, it may be a challenge to predict that the term is used in both of these meanings.

The fact that quite a few terms appear in the list underscores the key role that they play in legal English as an inherent aspect of its lexicon. Underlying this fact is the need to categorise different varieties and types of evidence. Although a general definition for evidence can be given, a particular item of evidence is perceived differently depending on whether or not it can be qualified by *oral*, *hearsay*, *sufficient*, *documentary*, etc. Most importantly, by classifying an item of evidence using one of these modifiers important legal implications follow. While in general English the need to classify evidence may not be acute, in legal English it seems to be a major driving force behind term formation.

Table 1 also includes a number of adjectives that serve either evaluative or descriptive functions, e.g. *fresh*, *new*, *cogent*, *live*, *credible*, *reliable*, *general*, *further*, and *available*. With the exception of *cogent*, *live*, *credible* and *general*, they all appear fairly frequently in the WBNC. It is worth noting in passing that the words *fresh* and *new* can be considered as (near)-synonyms, for example here: “If the fresh evidence were always evidence of primary fact, or new expert evidence, the test might be satisfactory” [2011] UKSC 18. Additionally, each of the words can potentially be subject to terminologisation, e.g. by being defined for the purposes of a given legal act. For instance, the Criminal Justice Act 2003, s. 78(2) states that: “Evidence is new if it was not adduced in the proceedings in which the person was acquitted (nor, if

those were appeal proceedings, in earlier proceedings to which the appeal related)”, thus severely restricting the meaning of *new* when used before *evidence*.

In the next step, the study also looked at the most underrepresented collocations of the noun *evidence* in the UKSCC as opposed to the WBNC. Given its restricted nature, it was natural to assume that a number of collocations would be overrepresented in the UKSCC. Yet, a number of collocations received a keyness score below 1, suggesting that they might constitute cases of underrepresentation. The data on the 39 collocates with the lowest keyness scores have been summarized in Table 2.

Table 2

The most underrepresented collocates of *evidence*
in the UKSCC as opposed to the WBNC

No	Collocate	Raw frequency in the WBNC	Raw frequency in the UKSCC	Normalised frequency in the WBNC	Normalised frequency in the UKSCC	Keyness score
1	Little	358	10	3.56	1.05	0.45
2	historical	84	0	0.84	0.00	0.54
3	archaeological	74	0	0.74	0.00	0.58
4	Research	70	0	0.70	0.00	0.59
5	considerable	81	1	0.81	0.11	0.61
6	first	54	0	0.54	0.00	0.65
7	early	50	1	0.50	0.11	0.74
8	more	118	6	1.17	0.63	0.75
9	clinical	33	0	0.33	0.00	0.75
10	contemporary	27	0	0.27	0.00	0.79
11	indirect	39	1	0.39	0.11	0.80
12	physical	22	0	0.22	0.00	0.82
13	much	112	7	1.11	0.74	0.82

14	less	21	0	0.21	0.00	0.83
15	fossil	19	0	0.19	0.00	0.84
16	recent	68	4	0.68	0.42	0.85
17	scant	16	0	0.16	0.00	0.86
18	definite	15	0	0.15	0.00	0.87
19	internal	14	0	0.14	0.00	0.88
20	overwhel- ming	37	2	0.37	0.21	0.88
21	visible	13	0	0.13	0.00	0.89
22	sound	13	0	0.13	0.00	0.89
23	abundant	36	2	0.36	0.21	0.89
24	vital	24	1	0.24	0.11	0.89
25	flimsy	12	0	0.12	0.00	0.89
26	molecular	11	0	0.11	0.00	0.90
27	current	22	1	0.22	0.11	0.91
28	wide- spread	10	0	0.10	0.00	0.91
29	reasona- ble	10	0	0.10	0.00	0.91
30	labora- tory	10	0	0.10	0.00	0.91
31	geological	10	0	0.10	0.00	0.91
32	dramatic	10	0	0.10	0.00	0.91
33	previous	9	0	0.09	0.00	0.92
34	irrefuta- ble	9	0	0.09	0.00	0.92
35	observa- tional	9	0	0.09	0.00	0.92
36	consi- sistent	9	0	0.09	0.00	0.92
37	computer	9	0	0.09	0.00	0.92
38	firm	20	1	0.20	0.11	0.92
39	hard	100	8	0.99	0.84	0.92

The first observation that can be made about Table 2 is that most of the scores are not particularly low, especially in comparison with the previous table which contained the data for the

most overrepresented collocations. In fact, if the same keyness threshold had been applied to the analysis of both the most overrepresented and the most underrepresented collocations, the table would include a single record for *little*. The reason for this lies in the nature of the WBNC and its size. As a general corpus, the WBNC comprises a wide range of texts encompassing a multitude of topics. It can therefore be expected to include a large pool of collocations (including legal collocations) that reflects its diverse composition. By contrast, the legal corpus can give prominence to a smaller group of collocations which are frequently employed in it.⁷ As a result and partly due to the size of the WBNC, for a collocation to meet the keyness threshold of 0.5, it would have to, for instance, appear at least 101 times in the WBNC and not a single time in the UKSCC. Such frequency requirements were considered too strict. Since the keyness score was employed only as a helpful tool for identifying cases of over- and underrepresentation, it was decided to discuss a similar number of collocates in both cases and, especially in the case of underrepresentation, place particular focus on general trends rather than individual examples.

Table 2 includes a small number of grammatical words that are common in general language, yet in the UKSCC are infrequent collocates of *evidence*. All of them are determiners: *much*, *more*, *little*, *less*, and *first*. Additionally, the first four can also be classified as quantifiers (Biber et al. 1999: 278). *Little* has the lowest keyness score (0.45) but the scores for the remaining four are also low. This is rather surprising. One might expect – given the high frequency of these words in general English – that the same pattern would also occur in the UKSCC. Yet, only sporadic instances of *much/more/little/less/first + evidence* are attested in the UKSCC. Closely related to *much* and *little*, albeit by no means synonymous, are the adjectives *sufficient* and *insufficient*, which by contrast occur in the UKSCC with a markedly higher

⁷ A similar explanation is offered in the study by L'Homme and Azoulay (2020: 160-161), who used a specialised corpus on the theme of the environment.

frequency than in the WBNC. This fact could be interpreted as testament to the role that the concept of (in)sufficient evidence plays in the legal system, which also justifies the inclusion of *insufficient evidence* as an entry in the *Longman Dictionary of Law* (Richards and Curzon 2011). The high keyness scores for *sufficient/insufficient* evidence on the one hand, and low ones for *little/much* evidence on the other, suggests that sufficiency of evidence is accorded greater importance in law than its mere quantity.

As regards lexical collocates that are featured in Table 2, several of them are clearly linked to specific fields other than law, e.g. *historical, archaeological, research, clinical, fossil, molecular, laboratory, and geological*. *Historical* is the most underrepresented in the UKSCC, occurring 84 times in the WNBC while not being attested in the UKSCC at all. The low frequency of collocations of *evidence* with these words is to be expected in view of the fact that *historical, archaeological, fossil* etc. *evidence* is unlikely to be adduced in court. Undoubtedly important in the fields to which they share ties, these collocations in the WBNC convey a sense of *evidence* that is closer to sense 1 in the definition quoted earlier.

Another group of collocates that can be distinguished on the basis of Table 2, includes *early, contemporary, recent, internal, visible, sound, current, widespread, previous, and computer*. These are of a largely descriptive character and the attributes that they denote appear – on the basis of the corpus evidence – to be more relevant in contexts other than law. Take the word *recent*, it is typically used with *evidence* in the WNBC in medical texts, as illustrated by the following sentence: “Recent evidence has shown that the expression of enzymes participating in biotransformation may play a part in tumour” (BNC). It thus places evidence at some point in time, while such chronological ordering may be considered not to be key in law.

The table also includes examples of collocates that are evaluative in character: *considerable, scant, definite, overwhelming, abundant, vital, flimsy, reasonable, dramatic, irrefutable, consistent, firm, and hard*. Some of them may be felt to be rather

strong and emotional (e.g. *flimsy*, *dramatic*, *irrefutable*, *definite*), which may make them unlikely collocates of *evidence* in a UKSC judgment. More importantly, they may seem to be less closely linked to important characteristics of evidence as seen in law. A good case in point is provided by *abundant*, whose meaning could be expressed as “existing in large quantities; more than enough” (OLD 2021), pointing to the fact that it is akin to *sufficient* but stronger. Based on its meaning one could expect *abundant* to collocate with *evidence* in the UKSCC. Stylistically, the adjective is rather formal and for this reason too it would blend in well with the rather formal style employed in judgments. Nevertheless, it is *sufficient evidence* that occurs with a markedly higher frequency than *abundant evidence*. A reason for this could be the fact that sufficient evidence is related to what is known as the evidential burden. As noted by Keane and McKeow (2012: 37): “A party discharges an evidential burden borne by him by adducing sufficient evidence for the issue in question to be submitted to the jury (tribunal of fact)”. Since the requirement mentioned in the quote involves only sufficient rather than abundant evidence, qualifying evidence as abundant might appear superfluous to a certain measure and also rather removed from the legal requirements that are in place. That is not to say that *abundant evidence* is an impossible combination in a UKSC judgment, but merely that it is unlikely as the corpus evidence from the UKSCC bears out. One can only agree with Sinclair, when he states that there “are virtually no impossible collocations, but some are much more likely than others” (1996: 411). It should be noted here that *abundant* is not the only adjective that has an analogue in Table 1. *Scant* expresses an idea roughly similar to *insufficient*, while *firm* and *hard* are similar to *compelling*, and last but not least *definite* and *irrefutable* are close to *conclusive*. The preference for certain collocates over others with a closely related meaning (as evidenced by *evidence* in the UKSCC) points to the conventional nature of collocations in legal English.

To achieve its third aim, a word sketch was generated for *evidence* in both the UKSCC and the WBNC and the candidate

collocates were sorted by frequency. The following discussion will be limited to the 50 most frequent collocates in each corpus. It is plausible to assume that dictionary compilers may want to focus on the frequency of a given combination as an important criterion for the selection of collocations. Since space is at a premium in many dictionaries and will prevent many dictionary compilers from including more than about 20 collocations, for the modifier + word category, a list of the top 50 candidate collocates provides a sound foundation for rank comparisons.

Table 3 displays the first fifty collocates of *evidence* in the UKSCC (listed in the second column) and the WBNC (listed in the third column). The first column indicates the rank of a collocate in the frequency lists obtained from each corpus. The third column, then, for each collocate obtained from the UKSCC, gives the rank assigned to it in the top 50 collocates obtained from the WBNC. To indicate that a given collocate was not among the top 50 collocates of evidence obtained from the WBNC, “n/a” has been used.

Table 3
Comparison of the ranks of collocates
in the UKSCC and the WBNC

Rank	Collocates in the UKSCC (sorted by frequency)	Rank of the collocate in the top 50 collocates in the WBNC	Collocates in the WBNC (sorted by frequency)
1	expert	27	further
2	oral	21	clear
3	fresh	29	new
4	new	3	empirical
5	relevant	49	sufficient
6	medical	10	good
7	further	1	direct
8	hearsay	44	available
9	sufficient	5	strong
10	documentary	12	medical

11	conclusive	14	scientific
12	available	8	documentary
13	cogent	n/a	hard
14	direct	7	conclusive
15	good	6	ample
16	clear	2	historical
17	reliable	48	considerable
18	statistical	42	circumstantial
19	additional	43	experimental
20	admissible	n/a	archaeological
21	prima facie	45	oral
22	insufficient	22	insufficient
23	objective	37	research
24	live	n/a	recent
25	credible	n/a	anecdotal
26	criminal	n/a	forensic
27	anonymous	n/a	expert
28	factual	n/a	convincing
29	false	n/a	fresh
30	inadmissible	n/a	real
31	opinion	n/a	early
32	material	n/a	independent
33	written	39	detailed
34	strong	9	indirect
35	unchallenged	n/a	overwhelming
36	ample	15	abundant
37	compelling	n/a	objective
38	empirical	4	substantial
39	character	n/a	written
40	detailed	n/a	conflicting
41	scientific	n/a	clinical
42	general	n/a	statistical
43	DNA	n/a	additional
44	primary	n/a	hearsay
45	conflicting	n/a	prima facie

46	convincing	28	positive
47	decisive	n/a	contemporary
48	experimental	19	reliable
49	post-published	n/a	relevant
50	specific	n/a	vital

Even a short glance at the table presented above indicates that the lists of collocations retrieved from the two corpora are strikingly different. This disparity was amplified in part as a result of removing some candidate collocates with the lowest informative value (e.g. *such*) from the list.

The overall picture that emerges from comparing the top 10 collocates retrieved from the UKSCC and the WBNC is one of vast discrepancies between them. Of the top 10 of the top 50 collocates of *evidence* in Table 3, only 3 are shared by both the legal and the general corpora (*new*, *further* and *sufficient*). In fact, it might be difficult to resist the impression that the lists of top 10 collocates in the UKSCC and the WBNC in fact correspond to two different nodes, rather than the same one but in different corpora. If the top 20 collocations are considered, the number of shared collocates grows to 9, achieving its highest overlap rate of 45 %. Expanding the list to the first 30, 40, and 50 collocates, results in a very slight increase in the number of shared collocates, amounting to 10, 14 and 16 respectively, and in a gradual decline in the percentage of overlapping collocates (33 %, 35 %, and 32 % respectively). As many as 23 collocates from the top 50 collocate list obtained from the UKSCC did not make the corresponding top 50 list of collocates based on the WBNC. An additional comparison between the ranks of the collocates in the two lists adds further weight to the claim regarding the wide discrepancies between them. The word *expert* ranks first on the UKSCC list but only 27th on the WBNC list. *Fresh* ranks 3rd and 29th respectively. *Relevant* is also an extreme case, ranking 5th and 49th. These are but a few examples. Many others can be found in the table. One conclusion that offers itself up from the comparison is that the potential of each corpus can be harnessed to produce lists of collocations with *evidence*,

yet the results will be considerably affected by the differences between the corpora and the language they are intended to represent.

5. Concluding remarks

The analysis reported in the previous section set out to establish and explore the most over- and underrepresented modifier + noun collocations of *evidence* in a legal corpus as opposed to a general corpus, and investigate the potential of the two corpora for the retrieval of such collocations. The main findings can be summarised as follows:

(1) The legal corpus was found to exhibit a strongly marked tendency to overrepresent certain collocations. As could be expected, a large subset of them was constituted by combinations that displayed various degrees of terminological character. However, a not insignificant number of combinations were identified whose overrepresentation could not be put down to their status as terms. Three of them were classified as (semi-)determiners. As for the remaining ones, it was suggested that what could better explain the prominence of such collocations in the corpus is the particular relevance of the attributes that the modifiers of *evidence* acquire when used in legal contexts.

(2) The legal corpus was found to reveal a slight-to-moderate tendency to underrepresent certain collocations, indicating a large pool of collocates of *evidence* that are shared by the two corpora and a somewhat limited systematic avoidance of a small group of collocates exhibited by the legal corpus. Apart from a significant proportion of collocates whose underrepresentation could easily be accounted for by pointing to the close link that they share to fields other than law, the study also identified 5 determiners and a large number of other collocates whose underrepresentation could be explained by their limited relevance in legal contexts or by the conventional nature of certain collocations, which results in a preference for particular collocates over others with a similar meaning.

(3) While both corpora were found to provide fertile material for collocation retrieval, a marked divergence was noted between the lists of collocations of *evidence* obtained from each of them. Many collocates were assigned vastly different ranks in the lists and the overlap between the top 50 collocations from both corpora stood at just 32 %. This figure is strikingly close to the percentage reported by L'Homme and Azoulay (2020: 166), despite the differences in scope and methodology between their study and the current one.

Based on these findings, several tentative observations can be made. Even for a word that is frequently used both in legal (specialised) and in general language and has a relatively simple sense structure, its collocational profile can vary sharply between a general and a legal (specialised) corpus. This might explain why a high level of competence in general language does not automatically translate into an equally high level of competence in legal (specialised) language and vice versa, which is a phenomenon that has been observed, among others, by Frankenberg-Garcia (2018) and Giczela-Pastwa (2021: 191). Due to the extent of the differences in collocations retrieved from a general corpus and a legal one, linguistic resources offering collocations compiled from a general corpus may prove to be of limited use when the context calls for legal (specialised) collocations and vice versa. As a consequence, this may warrant the need to develop such resources and language materials that target collocations in specific LSPs rather than in general language. Despite the insights produced into the nature of collocations, the study is not without limitations. It has to be acknowledged that the choice of corpora to represent general and legal language may have exerted some influence on the number and diversity of the collocations retrieved from them. The written component of the general corpus covers texts that were produced earlier than those in the legal one. Additionally, the UKSCC contains only UKSC judgments spanning a limited time frame and cannot thus capture the full complexity and wide diversity of legal English as used in the UK. Coupled with the study's focus on collocations featuring a single term and, to boot, only on the modifier + noun type, all these facts limit the generalisability of the

findings, although their combined impact is difficult to estimate. It is hoped that future investigations will address some of these limitations since much further research into the nature of legal (specialised) collocations is needed.

References

- [2011] UKSC 18 = R (on the application of Adams) v Secretary of State for Justice (2011). United Kingdom Supreme Court, case 18. Available at <<http://www.bailii.org/uk/cases/UKSC/2011/18.html>>. Accessed 15.06.2021.
- [2017] UKSC 21 = BPE Solicitors and another v Hughes-Holland (in substitution for Gabriel) (2017). United Kingdom Supreme Court, case 21. Available at <<http://www.bailii.org/uk/cases/UKSC/2017/21.html>>. Accessed 15.06.2021.
- Criminal Justice Act 2003*. Available at <<https://www.legislation.gov.uk/ukpga/2003/44>>. Accessed 15.06.2021.
- Badziński, Arkadiusz (2019). "Problems in medical translation among professional and non-professional translators: Collocations as a key issue". *Beyond Philology* 16/4: 157–177.
- Baker, Paul, Andrew Hardie, Tony McEnery (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Benson, Morton (1989). "The structure of the collocational dictionary". *International Journal of Lexicography* 2/1: 1–14.
- Bergenholtz, Henning, Sven Tarp (1994). "Mehrworttermini und Kollokationen in Fachwörterbüchern". In: Burkhard Schaefer, Henning Bergenholtz (eds.). *Fachlexikographie: Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr Verlag, 385–419.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Biel, Łucja (2010). "Corpus-based studies of legal language for translation purposes: Methodological and practical potential". In: Carmen Heine, Jan Engberg (eds.). *Reconceptualizing LSP: Online Proceedings of the XVII European LSP Symposium 2009*. Aarhus: Aarhus University Aarhus, 1–15.
- Biel, Łucja (2012). "Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law". In:

- Antonio Pamies, José Manuel Pazos Breaña, Lucía Luque Nadal (eds.). *Phraseology and Discourse: Cross Linguistic and Corpus-based Approaches*. Baltmannsweiler: Schneider Verlag, 225–233.
- Biel, Łucja (2014). “Phraseology in Legal Translation: a Corpus-based Analysis of Textual Mapping in EU Law”. In: Le Cheng, King Kui Sin, Anne Wagner (eds.). *Ashgate Handbook of Legal Translation*. Surrey: Ashgate Publishing.
- BNC 2015 = What is the BNC? Available at <<http://www.natcorp.ox.ac.uk/corpus/index.xml>>. Accessed 15.06.2021.
- Boers, Frank, Stuart Webb (2017). “Teaching and learning collocation in adult second and foreign language learning”. *Language Teaching* 51/1: 77–89.
- Buendía Castro, Miriam, Pamela Faber (2015). “Phraseological units in English-Spanish legal dictionaries: A comparative study”. *Fachsprache* 37/3-4: 161–175.
- CD 2021 = *Cambridge Dictionary*. Available at <<https://dictionary.cambridge.org>>. Accessed 15.06.2021.
- Firth, John R. (1957). “Modes of meaning”. In: J. R. Firth. *Papers in Linguistics, 1934-1951*. London: Oxford University Press, 190–215.
- Firth, John R. (1968). “A synopsis of linguistic theory, 1930–1955”. In: F. R. Palmer (ed.). *Selected Papers of J. R. Firth 1952–1959*. Bloomington: Indiana University Press, 168–205.
- Frankenberg-Garcia, Ana (2018). “Investigating the collocations available to EAP writers”. *Journal of English for Academic Purposes* 35: 93–104.
- Garner, Bryan A. (1995). *A Dictionary of Modern legal Usage*. 2nd edition. Oxford: Oxford University Press.
- Garner, Bryan A. (ed.) (2004). *Black’s Law Dictionary*. 8th edition. St. Paul: Thomson West.
- Giczela-Pastwa, Justyna (2021). “Developing phraseological competence in L2 legal translator trainees: A proposal of a data mining technique applied in translation from an LLD into ELF”. *The Interpreter and Translator Trainer* 15/2: 187–204.
- Gledhill, Christopher J. (2000). *Collocations in Science Writing*. Tübingen: Gunter Narr Verlag.
- Goźdz-Roszkowski, Stanisław (2011). *Patterns of Linguistic Variation in American Legal English: A Corpus-Based Study*. Frankfurt am Main: Peter Lang.
- Goźdz-Roszkowski, Stanisław, Gianluca Pontrandolfo (2015). “Legal phraseology today: Corpus-based applications across legal languages and genres”. *Fachsprache* 37/3-4: 130–138.

- Heid, Ulrich (2001). "Collocations in sublanguage texts: Extraction from corpora". In: Sue Ellen Wright (ed.). *Handbook of Terminology Management: Applications-Oriented Terminology Management*. Vol. 2. Amsterdam – Philadelphia: John Benjamins Publishing Company, 788–808.
- Hoey, Michael (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Keane, Adrian, Paul McKeown (2012). *The Modern Law of Evidence*. 9th edition. Oxford: Oxford University Press.
- Kilgarriff, Adam (2005). "Language is never ever ever random". *Corpus Linguistics and Linguistic Theory* 1/2: 263–276.
- Kilgarriff, Adam (2009). "Simple maths for keywords". In: Michaela Mahlberg, Victorina González-Díaz, Catherine Smith (eds.). *Proceedings of the Corpus Linguistics Conference, CL2009, University of Liverpool, UK, 20–23 July 2009*. Liverpool: University of Liverpool. Available at <http://ucrel.lancs.ac.uk/publications/cl2009/171_FullPaper.doc>. Accessed 15.06.2021.
- Kjaer, Anne Lise (1990a). "Context-conditioned word combinations in legal language". *Terminology Science & Research* 1/1-2: 21–32.
- Kjaer, Anne Lise (1990b). "Phraseology research – State-of-the-art". *Terminology Science & Research* 1/1-2: 3–20.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel (2014). "The Sketch Engine: Ten years on". *Lexicography* 1: 7–36.
- Lexical Computing (2015). *Statistics used in the Sketch Engine*. Available at <<https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf>>. Accessed 15.06.2021.
- L'Homme, Marie-Claude (2000). "Understanding specialized lexical combinations". *Terminology* 6/1: 89–110.
- L'Homme, Marie-Claude, Daphnée Azoulay (2020). "Collecting collocations from general and specialised corpora: A comparative analysis". In: Gloria Corpas Pastor, Jean-Pierre Colson (eds.). *Computational Phraseology*. Amsterdam – Philadelphia: John Benjamins Publishing Company, 151–175.
- L'Homme, Marie-Claude, Patrick Leroyer (2009). "Combining the semantics of collocations with situation-driven search paths in specialized dictionaries". *Terminology* 15/2: 258–283.
- Łukasik, Marek (2017). "Lexical minimum (re)defined." *Lingwistyka Stosowana* 23/3: 47–63.

- Mair, Christian (2007). "Varieties of English around the world: Collocational and cultural profiles". In: Paul Skandera (ed.). *Phraseology and Culture in English*. Berlin: Mouton de Gruyter, 437–468.
- Martin, Willy (1992). "Remarks on collocations in sublanguages". *Terminologie et Traduction* 2-3: 157-164.
- Matilla, Heikki E. S. (2012). "Legal vocabulary". In: Peter M. Tiersma, Lawrence M. Solan (eds.). *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, 27–38.
- Meyer, Ingrid, Kristen Mackintosh (1994). "Phraseme analysis and concept analysis in exploring a symbiotic relationship in the specialized lexicon". In: Willy Martin, Willem Meijs, Margreet Moerland, Elsemie ten Pas, Piet van Sterkenburg, Piek Vossen (eds.). *Euralex 1994 Proceedings*. Amsterdam, 339–348.
- Michta, Tomasz (2007). "Zastosowanie narzędzi korpusowych do badania kolokacji w tekstach specjalistycznych na przykładzie angielskich tekstów chemicznych". In: Marek Łukasik (ed.). *Debiuty Naukowe I*. Warszawa: Katedra Języków Specjalistycznych UW, 55–61.
- Michta, Tomasz, Maria Kloza, Jan Łompieś, Mariusz Mela, Wioletta Mela, Magdalena Miąc, Joanna Newska, Liliana Religa (2009). "Studentki Słownik Kolokacji Angielskiego Języka Medycyny". In: Marek Łukasik (ed.): *Debiuty Naukowe III*. Warszawa: Katedra Języków Specjalistycznych UW, 89–225.
- Michta, Tomasz (2018). *A Model for an English-Polish Systematic Dictionary of Chemical Terminology*. Siedlce: Scientific Publishing House of Siedlce University of Natural Sciences and Humanities.
- Michta, Tomasz, Katarzyna Mroczyńska (2022). *Towards a Dictionary of Legal English Collocations*. Siedlce: Wydawnictwo Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach.
- Montero-Martínez, Silvia, Miriam Buendía-Castro (2012). "La sistematización en el tratamiento de las construcciones fraseológicas: el caso del medio ambiente". In: Izaskun Elorza, Ovidi Carbonell i Cortés, Reyes Albarrán, Blanca García Riaza, Miriam Pérez-Veneros (eds.). *Empiricism and Analytical Tools for 21st Century Applied Linguistics*. Salamanca: Ediciones de la Universidad de Salamanca. 711–724.
- Mroczyńska, Katarzyna (2020). "A dictionary of legal English collocations as an aid for mastering the legal English genre". *Linguistics Beyond and Within* 6: 130–141.
- Nesselhauf, Nadia (2004). *Collocations in a Learner Corpus*. Amsterdam – Philadelphia: John Benjamins Publishing Company.

- OLD 2021 = *Oxford Learner's Dictionaries*. Available at <<https://www.oxfordlearnersdictionaries.com>>. Accessed 15.06.2021.
- Picht, Heribert (1987). "Term and their LSP Environment – LSP Phraseology". *Meta* 32/2: 149–155.
- Rzepkowska, Agnieszka (2021). "Labour-law basic terminology based on the Polish Labour Code". *Półrocznik Językoznawczy Tertium* 6/1: 257–283.
- Richards, P. H., L. B. Curzon (2011). *The Longman Dictionary of Law*. 8th edition. Harlow: Longman.
- Saber, Anthony, Audrey Cartron, Claire Kloppmann-Lambert, Céline Louis (2020). "Towards a typology of linguistic and stylistic errors in scientific abstracts written by low-proficiency Geoscience and Mechanical Engineering doctoral students in France". *Fachsprache* 42/3-4: 90–114.
- Sinclair, John (1966). "Beginning the study of lexis". In: C. E. Bazell, J. C. Catford, M. A. K. Halliday, R. H. Robins (eds.). *In Memory of J. R. Firth*. London: Longman, 410–430.
- Sinclair, John (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Tiersma, Peter M. (2002). *Legal Language*. Chicago: University of Chicago Press.
- Tiersma, Peter M. (2008). "The nature of legal language". In: John Gibbons, M. Teresa Turell (eds.). *Dimensions of Forensic Linguistics*. Amsterdam – Philadelphia: John Benjamins Publishing Company: 7–25.
- UKSC 2021a = Role of the Supreme Court. Available at <<https://www.supremecourt.uk/about/role-of-the-supreme-court.html>>. Accessed 15.06.2021.
- UKSC 2021b = The Supreme Court. Available at <<https://www.supremecourt.uk/about/the-supreme-court.html>>. Accessed 15.06.2021.
- Ward, Jeremy (2007). "Collocation and technicality in EAP engineering". *Journal of English for Academic Purposes* 6: 18–35.

Tomasz Michta
ORCID iD: 0000-0001-8997-6154
Uniwersytet w Białymstoku
Wydział Filologiczny
Plac Niezależnego Zrzeszenia Studentów 1
15-420 Białystok
Poland
t.michta@uwb.edu.pl

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

<https://doi.org/10.26881/bp.2022.1.04>

Attempt to reconstruct the linguistic picture of a *true Pole*

IWONA BENENOWSKA

*Received 17.08.2021,
received in revised form 14.11.2021,
accepted 10.01.2022.*

Abstract

The attempt at reconstructing the contemporary image of a true Pole aims to obtain the basis for certain remarks resulting from the observation and interpretation of the linguistic phenomenon under study. The research paradigm of the linguistic picture of the world (LPW) was used. The sources of linguistic material were mainly the *National Corpus of the Polish Language* (in the period 1981-2020, with particular emphasis on the 20 years of the turn of the centuries) and selected Internet texts.

Keywords

linguistic picture of the world (LPW), a true Pole, *National Corpus of the Polish Language* (Polish abbrev. NKJP), reconstruction, features of a Pole

Próba rekonstrukcji językowego obrazu *prawdziwego Polaka*

Abstrakt

Celem tekstu jest próba zrekonstruowania obrazu *prawdziwego Polaka*, aby uzyskać podstawy do poczynienia pewnych uwag wynikających z obserwacji i interpretacji badanego zjawiska językowego. Wykorzystano paradygmat badawczy językowego obrazu świata (JOS). Źródło materiału językowego stanowił przede wszystkim *Narodowy Korpus Języka Polskiego* (w cezurze czasowej 1981-2020, ze szczególnym uwzględnieniem 20 lat przełomu wieków), a także wybrane teksty internetowe.

Słowa kluczowe

językowy obraz świata (JOS), prawdziwy Polak, *Narodowy Korpus Języka Polskiego* (NKJP), rekonstrukcja, cechy Polaka

The text aims to reconstruct the image of a *true Pole*, providing a basis for specific interpretations of the research outcomes. The research paradigm of the linguistic picture of the world (LPW) was used. The sources of linguistic material were mainly the *National Corpus of the Polish Language* (with examples from the years 1981-2020, with the predominance of the 20 years at the turn of the centuries), and Internet texts.

The *National Corpus of the Polish Language* (NKJP) is an indispensable source of linguistic material for researchers of contemporary Polish. Not only can linguists use it, but also specialists in other fields of humanities or computer science. Poles badly need this large linguistic corpus, balanced in terms of genre and subject matter, the online treasury of Polish (as the project's creators call it)¹. The reference corpus of Polish con-

¹ Other nations, such as the British, the Germans, the Czechs, and the Russians also have national corpora.

tains about one hundred thousand words. Due to their technical possibilities, Corpus search engines (for NKJP data – PELCRA) enable efficient browsing of text resources. For this purpose, they use advanced tools taking the variety of Polish words into account and analysing sentence structures. The possibilities of finding concordances, indicating collocations, choosing sorting or grouping options, narrowing the search to occurrences of matches in texts of a given functional type, title, or publication date turn out to be excellent facilitation for the researcher. Moreover, the selected example can always be detailed bibliographically, or the cited passage can be expanded (right menu), gaining a broader context. It allows isolating typical uses of words and constructions, studying their meaning and function, observing the language evolution, etc.²

1. Introduction

The word *Pole* is not recorded by all dictionaries, and they define it emphasising the origin and territorial affiliation:

- a person of Polish nationality, a resident of Poland, a citizen of that country (USJP)
- a person of Polish nationality
Usage notes: Depending on the circumstances of usage, the word may also mean a person having Polish citizenship or residing in Poland or even an athlete playing for a Polish team, regardless of his/her nationality.
subject qualification: names of persons regarding their origin and territorial affiliation (WSJP)
- someone who comes from Poland (ISJP)
- a citizen of Poland (SJP net. 6)
- a person born in Poland, a Polish citizen (SW net. 7)

² For more see: nkjp.uni.lodz.pl (accessed: 10.05.2021).

In many dictionaries, e.g., *Dictionary of the Polish Language* (*Słownik języka polskiego*) (SJPSz) edited by Mieczysław Szymczak, *Dictionary of Contemporary Polish Language* (*Słownik współczesnego języka polskiego*) (SWJP) edited by Bogusław Dunaj, *New Dictionary of the Polish Language* (*Nowy słownik języka polskiego*) (NSJP) edited by Elżbieta Sobol, *Popular Dictionary of Polish Language* (*Popularny słownik języka polskiego*) (PSJP) edited by the latter, *Practical Dictionary of Correct Polish Language Not Only for Young People* (*Praktyczny słownik poprawnej polszczyzny nie tylko dla młodzieży*) (PSPP) edited by Andrzej Markowski, there is no such entry. Maybe in line with the principle: a horse is a horse is a horse / what is the horse everyone sees (in Polish: Jaki jest koń każdy widzi).

Meanwhile, we know from history that in the First Republic of Poland, the term *a Pole/a free Pole* referred exclusively to the nobility as *Patris Patriae* (= *Ojcowie Ojczyzny*) (cf. Opaliński 1995: 344). The privileged position of the nobility in society was abolished only by the March Constitution of 1921 (Article 96). Nowadays, there are at least **two** possible approaches to the problem of defining the notion of a Pole as a member of the Polish nation: 1. the feeling of "being a Pole" may be expressed by using the Polish language and treating it as the mother tongue, observing customs characteristic of Polish culture, the conviction of having common Polish roots or a sense of Polish national identity:

A nation is a permanent community of people historically formed based on historical fate, culture, language, territory, and economic life, manifested in the national consciousness of its members (Rulka 2001: 21).

2. the Polish nation consists of all Polish citizens (Poles), who feel a common heritage and good of Poland, regardless of nationality. It does not narrow the concept of nation only to Polish citizenship:

In concern for the existence and future of our Homeland, having regained in 1989 the possibility of sovereign and democratic determination of its fate, we, the Polish Nation – all citizens of the Republic of Poland [...] (Preamble to the Constitution of the Republic of Poland 1997³).

Why is this mentioned? Because further on, statements concerning a true Pole invoke both territorial and social criteria.

If we add the exponent of the verbal modality of *true* to the word, the meaning is modified:

The modifier 'true' – just like 'normal', 'typical', 'ordinary', 'average'; it introduces the **subjectification** of judgment, refers to a **subject** who is the instance that establishes the **norm**. To each of such modifiers one can thus add a metatextual formula: "according to the speaker", "in the speaker's view", "in my opinion", "in his opinion" (Bartmiński, Panasiuk 1993: 370; emphasis by the author).

The choice of certain aspects and the ways they are filled in impact the cognitive structure of the expression. The Pole's characterisation becomes internally structured in a certain non-accidental way. After the initial categorisation ('a person who...'), overtly relational, subjectivising features come to the fore: 'proud', 'is a Catholic', 'is a patriot', etc., followed by characteristics concerning function, location, objects used, and others.

In his/her self-characteristics (the ASA'90⁴ questionnaire), a Pole emphasises three "canonical" features: patriotism, courage, and

³ The cited Constitution was passed on 2 April 1997, approved in a national referendum on 25 May 1997, and came into force on 17 October 1997. The Constitution is the most important legal act in Poland, consisting of a preamble and 13 chapters, divided into 243 articles (see Preambuła – Konstytucja Rzeczypospolitej Polskiej [Konst.] (arslege.pl) (accessed: 10.06.2021)).

⁴ Materials from a survey carried out as part of the research topic *Słownictwo aksjologiczne języka polskiego (Axiological vocabulary of Polish)*, 1990, held by the Department of Textology and Grammar of Modern Polish Language of Maria Curie-Skłodowska University in Lublin.

ASA – a survey of the axiological dictionary, materials collected among students of Lublin state universities in 1990, 2000, and 2010. The respondents answered the question: What, in your opinion, constitutes the essence of the "real him" X?, where in each survey the names of values were substituted for

hospitality, in order also: honour, pride, attachment to freedom, romanticism. Moreover, contemporary Poles are characterised as a "*drunkard*" and "*peddler*", and – concerning their attitude to work – either hard-working or lazy. [...] It is worth adding, by the way, that this auto-stereotype contrasts *in plus* with hetero-stereotypes of Poles functioning abroad (e.g., for the French a Pole is 'drunk' and 'dirty', for the Czechs – 'pathetic', for the Germans – 'nationalistic' and 'unmanageable', for the Russians – 'Catholic', 'hypocritical' and 'rebellious') (Bartmiński, Panasiuk 1993: 380).

2. Characteristics of a true Pole

What image does the expression 'a true Pole' generate? Let us trace this image created in the texts of 1980s – 2020s, collected especially in the National Corpus of the Polish Language (NKJP) and other Internet sources (henceforth net.).

From the Poles' point of view (self-created image), the image of a true Pole includes such features as:

2.1. a Catholic: *Because a true Pole will always be in some sense a **Catholic**.*⁵ [...] (NKJP 1); *A real **Catholic Pole**, on the other hand, has never accepted that there really were the Jews [...]* (NKJP 4); *We have a national **Catholicism**. These people think that a real Pole is only a Pole-Catholic who votes for their party. [...]* (NKJP 6); [...] *1989, when the Civic Committees coined the slogan that a real Pole is a **Catholic**.* (NKJP 8);⁶ *The conviction*

X from the list of 100 entries selected for the axiological dictionary. For research results, see, e.g., *Język – wartości – polityka. Zmiany rozumienia nazw wartości w okresie transformacji ustrojowej w Polsce. Raport z badań empirycznych (Language – values – politics. Changes in the understanding of the names of values during political transformation in Poland. Report from empirical research.)* (Bartmiński 2006). See also the report and the list of entries – Małgorzata Mazurkiewicz (1991: 257-264).

⁵ In all quoted passages, emphasis by the author of the article.

⁶ The next sentence shows the dangers of such an attribution: *So, most inhabitants of Cieszyn Silesia immediately felt themselves to be second-class citizens* – the exclusion of adherents of other faiths. This is particularly noticeable in the attitude of Poles-Catholics towards the Jews. Adam Szostkiewicz writes about it in the article *Jeżus by się dziwił, co wyprawiamy* (*Jesus would be surprised, what we are doing*) (NKJP 11), devoted, among other things, to the activities of Father Stanisław Musiał, an advocate of Catholic-Jewish

of one's own uniqueness is part of the identity of a **Pole-Catholic**. It is not enough to be **a Pole and a Catholic**, one must be a real Pole and a real **Catholic**. (NKJP 11); A TRUE POLE – is in principle like a **Catholic Pole**, but must have some additional features. (NKJP 21); 90 % of the Poles are **Catholics** (net. 2); [a true Pole] Should be **a Catholic**. (net. 3); **Catholic Poland**. Now mandatory for every "good Pole". [Kaczynski:] Christianity is part of our national identity. **The Church** was and is the preacher and holder of the only universally known system of values in Poland." (net. 12); The eternal call of "real Poles, real patriots" is of course "**God, honour, homeland**". And this is where the term "**God-fearing patriotism**" comes from. In this classical triad, we also have a simple equal sign by implication: a real Pole (or elsewhere: a real Irishman, Spaniard, etc.) can only be someone who believes in **God**. (net. 14); Right-wing religiosity in our country is understood exclusively as **Catholicism**. [...] A real Pole is **a Catholic** and that does it! (net. 15);

– a defender of the Church and Christian values: And finally, a special candidate: [...] **defends the Church** [...] – in a word a 'true' Pole. (NKJP 2); You spread such filth into the world that it hurts the heart. (This is not how a real Pole acts!) In the name of the Radio Maryja Family we PROTEST VEHEMENTLY against publicly **insulting our holiness**. [...] (NKJP 10); True Poles are already jumping on me. You disrespect the Roman Catholic religion! It was his **deep faith** that made Lech reject communism and become an oppositionist, a Christian one by the way, and not a secular one like those KOR members. (NKJP 13); [...] in my opinion, this "true Pole" is the one who **listens to the priest** every Sunday and instead of using his/her reason and simply thinking on his/her own, **lets himself/herself be led** like silly lambs by his/her shepherd... (net. 18);

– cooperates with the hierarchs of the Church, rather than seeking conflict with them: [...] **the right is on Rydzyk's side**. "President Kaczyński has finally revealed who he represents. – we

cooperation, a member of the Polish Episcopal Commission for Dialogue with Judaism since 1986.

read in one of the letters to the editor. – First, **he liquidated the best** archbishop Stanisław Wielgus and now he **is destroying Radio Maryja and Television Trwam** headed by Fr Rydzyk. Is there really no true Pole for President in Poland?" (NKJP 12);

– is opposed to the idea of evolutionism: Well, perhaps Mr Kroll, like his prophet Darwin, has discovered monkey-like qualities in himself. If so, I am even happy to agree with that. But I see no reason why a true Pole and Catholic **should admit to these genes**. (NKJP 9)

– trusts the Catholic media: A true Pole is a **faithful listener of Radio Maryja and reader of "Gazeta Polska"**. (net. 1);

– a role model or authority in this respect are for him/her:

Lech Wałęsa: We vote for **Wałęsa**, he is a true Pole and a Catholic! – I heard. (NKJP 3a⁷); Mr President **Lech Wałęsa** is our only Polish saviour. [...] A true Pole, a Catholic⁸ [...] (NKJP 5);

Fr. Tadeusz Rydzyk (a founder and director of Radio Maryja, Television Trwam), Bishop Edward Frankowski (a chaplain of "Solidarity" (NSZZ "Solidarność") in 1980– 1992), Saint John Paul II: **How dare you insult and undermine such authorities as Fr. Rydzyk, bishop Frankowski, not to mention the greatest of Poles – John Paul II**. (NKJP 10)

Fr. Henryk Jankowski (a chaplain of "Solidarity", removed from the post of parish priest St Brygida parish in Gdańsk in 2004): **Because "true Poles", of course, wait only for his** [prelate Jankowski's] speeches. (NKJP 14); **Priest Jankowski** is a true Pole. **We want the President to be purely from Poland**", said the former employee of "Star" (NKJP 16);⁹

⁷ The subscripts *a, b*, etc. indicate that these are Internet users' consecutive statements on the same discussion thread.

⁸ The notation of all quoted examples follows the original ones.

⁹ It is worth noting at this point that the analysed texts also point to models of "a true Pole", e.g., Marek Edelman: Havel invariably holds **Edelman** in high esteem. When he learns that his biography is being written, he immediately writes him a short letter: "I deeply respect all that you have accomplished in your life, your attitude, your courage, and I am happy that I had the opportunity to meet you personally. You are for me **an example of a true Pole**, an authentic embodiment of what is best in Poland". (NKJP 70); Wojciech Cejrowski: **Wojciech Cejrowski** has his programme on one of the radio stations [...]. Once Mr Wojtek from Siedlce called in and was happy on air: "**I am happy that I am talking to a true Pole**". (NKJP 72).

2.2. feels strong ties with his nation (national views, sense of national identity¹⁰): Yes, a true Pole always walks and **must walk a straight and national path**. (NKJP 15); After all, I had Bohdan, who **did everything with Poland in mind**. He was truly a true Pole. [...] **I was everywhere where Poland had to be mentioned**. (NKJP 16); [Paweł] Niezgodzki is counted among the so-called "true Poles", a **nationalist current among the activists of the** ["Mazowsze"] **region**. (NKJP 18); The canon of righteous conduct in Poland is changing. Until recently, an important part of it was **the attitude to the "Polish issue"**. Today it is not necessarily so. (NKJP 19); Bodo "did not hide his right-wing sympathies. He subscribed to the weekly »Prosto z mostu«¹¹ [...] he **attended** [as befits a True Pole] **Roman Dmowski's funeral**¹² at Bródno Cemetery". (NKJP 20); So, who will the Poles be in ten, twenty years? When they can no longer be [...] **nationalists and, together with Dmowski, turn away from foreigners**, [...]? (NKJP 22); A true Pole is a person who is not ashamed of his country and when travelling abroad **does not pretend to be of a different nationality**. He/she does notice the faults of his/her own country, nation, but only **to draw attention to the problems**. Moreover, he/she simply **demonstrates a civic attitude** – he/she tries on his/her own, to the extent that he/she is able, to improve something in his/her immediate vicinity, his/her environment. (net. 4); He is convinced

There are also examples of anti-patterns, e.g., Jan Kobylanski: [...] He is the most **pompous** "true Pole" south of the Rio Grande. [...] He suffers from a mania for grandeur, questions the decisions of successive Polish authorities, undermines the competence of Polish ambassadors, fights with Polish missions, intrigues [...]. (NKJP 71).

¹⁰ See in USJP: nationality – belonging to a particular nation or a designation of an ethnic and cultural community; also: a sense of belonging; national – pertaining to a nation, the whole population of a certain territory, characteristic of a particular nation, proper to it, belonging to it; identity – in relation to a community: an awareness of common features and a sense of unity.

¹¹ "Prosto z Mostu" was published between 1931 and 1935 as a Sunday cultural supplement to the ONR [Obóz Narodowo-Radykalny (National Radical Camp) – author's note] affiliated daily "ABC".

¹² Roman Dmowski was a co-founder of the National Democracy (in other words: Endecja, national movement), considered to be the main ideologist of the Polish nationalism.

that the Poles are **a noble nation**. (net. 1); He/she is a Pole who **is not ashamed of his/her country, who has a sense of collective responsibility** for himself/herself and for his/her compatriots, dead, present and future. (net. 3a); [...] a true Pole **is proud** to be a Pole. [...] by his/her behaviour he/she **does not bring shame** on Poland (he/she does not steal from the state). He/she does not have to be an idealist, but he/she should know the **Polish anthem**, know the **dates of national holidays**, and **respect the national flag**. (net. 3b); [...] according to ideologists of the Macierewicz type, a **true Pole** is someone who **continues the traditions of old nationalism**. (net. 13); We want to live in Poland, where Poles live, feel Polish, are Poles in flesh and blood. [...] and **they honour the traditions of 1918, which were instilled by Piłsudski, Roman Dmowski, Ignacy Jan Paderewski**, and the communists forget about it. (net. 19);

2.3. feels a patriot,¹³ is ready to sacrifice himself/herself for the homeland, fight for its freedom and good: [...] **every true Pole must be a soldier. Faithfulness** to the tormented homeland, "at any time and on any day". [...]. [For the last two years, he has been teaching us **love for our homeland. He instilled, propagated, and strengthened Polishness in us** [...]. (NKJP 24); Waldemar Lau does not feel like a hero, but only a true Pole and **patriot**. Those days nevertheless had a great influence on his further life. He lost a lot then: studies, good job, and health, but he notes only profits: recognition of those with whom he had to work, respect and a clear conscience fulfilled. (NKJP 25); He was truly a true Pole. [...] From 1974 until the end of his life, he was secretary of the AK Circle in Belgium. He **was active in many independence organisations**. At one time, he was even **a delegate of the Polish Government** in London for the Benelux countries. He was **President of the Association of Polish Veterans** in Belgium from 1988. (NKJP 17); He/she may even have

¹³ According to the definition of a patriot and patriotism in USJP: a patriot – one who loves his/her homeland and nation, ready to sacrifice for their sake; patriotism – love for the homeland, one's own nation combined with the readiness to bear sacrifices for them.

quite a large fortune, but **must declare his/her willingness to sacrifice everything he/she has for the good of his/her homeland** when the need arises.¹⁴ (NKJP 21); These elements of internal emigration, which we observe today, are not new in Poland. They also existed back then, and on an even greater scale. For example, one could go to a hospital to treat people, but not to the Ministry of Health for an administrative post, oh no!¹⁵ Because it's not appropriate, it's not in good taste, it's unpatriotic, it doesn't suit a true Pole! (NKJP 26); The most **common variety of patriotism** assumes **unconditional love for the homeland**, which is **understood sentimentally** as a familiar landscape, original cuisine, folk customs, etc., and finally the place where our ancestors were born and lived. The sentiment is guarded by the collective of true Poles, i.e., the ethnic element. [...] However, the **pro-state patriotism**, which has a very short history in Poland, is more interesting. Representatives of this trend do not see Polishness in the graves of their ancestors, in landscapes on the Vistula or in smelling bigos (sauerkraut), but, for example... **in paying increasing taxes honestly or in paying for the bus ticket that gets more expensive every year.** (net. 15);

– opposes socialism/communism, Nazism, liberalism, etc: A true Pole, a patriot should f...ck* **Nazism** and **communism**¹⁶ [NKJP 28]; **Piłsudski was unacceptable** for the right wing. **A man with a socialist past**, desiring instead of incorporation an equal union with Lithuania and Ukraine, accused of contacts with Freemasonry and religiously indifferent – could there be anything more repugnant for "a true Pole"? (NKJP 54); [True Poles] **Speak the language of radicals, say: 'lustration', 'de-communisation', do not want a market economy.** (NKJP 6); The oft-used **liberalism**, in the connotations of the past party, **meant something extremely indecent, shameful, very, very shameful,**

¹⁴ The next sentence, however, undermines the sincerity of the action, offering the possibility of free choice: *Has the need already arisen or not – he/she determines for himself/herself and does not suffer anyone to urge him/her to such a decision.*

¹⁵ It is about not cooperating with the communist government.

¹⁶ This quote also indicates intermediate traits: anti-communist and anti-Nazi.

which does not befit a true Pole. (NKJP 55); *On the one hand there was the [communist] government, which disgusted me, to which I had the worst possible attitude, from which I expected nothing good. On the other hand [Solidarity], a populist **wave** of ethnic emotions and **cave anti-communism.** [NKJP 49]; I saw one difference – the attitude to communism. "True Poles" from Solidarity also **represented totalitarian ideology, only in a colour other than red.** [...] the so-called **liberal circles**, whose core was the old opposition, **were squeezed out of Solidarity.** (NKJP 57); *The first element is the dichotomous division of the world into 'us' and 'you'. When "we" were progressive and fought for peace and "you" were the dwarfs of reaction and imperialists. Today "we" are the true Poles and "you" are liberals, enemies of Poland or even Europeans.* (NKJP 58); *We want to live in Poland, where the Poles live, feel Polish, are Poles in flesh and blood. [...]... and they honour the traditions of 1918, which were instilled by Piłsudski, Roman Dmowski and Ignacy Jan Paderewski, and **the communists forget about that.** [...] [The communist] Komorowski said that there will be a monument in Ujazdowskie Avenue, probably to **Mazowiecki**, you will see. **The man who betrayed Poland at the Round Table** (net. 19);**

- has a conviction that Poland is threatened by an enemy: [the Kaczyński brothers] *opened the presidential campaign in the National Museum against the background of the 'Battle of Grunwald'. The message was clear: it is we who are taking over the legacy of the victorious **fight against the eternal enemy.** (NKJP 56); [...] **the need for the constant presence of the enemy.** Once it was obviously a class enemy, now it is the enemy of the Polish nation or the enemy of the Church. (NKJP 58); [...] and 'you' are **liberals, enemies of Poland** or even Europeans. (NKJP 58); [...] *Russia and Germany, always hostile to us, and of course the Jews and Freemasons. [...] [Poland] is an unhappy nation only because of its **neighbours**, who are eternal **enemies**, and the **wickedness of allies**, who always treacherously abandon us in our time of need.* (net. 1);*
- is proud of Poland's historical role, giving the Poles special credit for saving the world: [is convinced that] *In the history of*

*the world, it is Poland that **has** always **saved it** [the world] **from perdition**. (net. 1a); On account of their **services to Humanity and having saved Europe from perdition so many times**, all True Poles should be granted special privileges and honours everywhere abroad. (net. 1b); The need to abstain from **displaying national pride** gives rise to frustration and subconscious resentment towards the whole world. (net. 1c);*

2.4. speaks Polish, recognizes this language as native: *So, call your language whatever you want, and say: "I am Polish, so **I speak Polish**". Worse if you say: "I am a real true Pole, so I give in my pure as tears language obvious proof that I am better than others". [...].¹⁷ (NKJP 23); A true Pole is one who [...] Also **knows his/her own language very well and follows its rules** of spelling or grammar. (net. 4); And I will say more, a true Pole [...] **cares about the purity of his/her language** [...]¹⁸ (net. 3);*

2.5. appreciates Polish culture, history, cultivates native customs, etc: *So, who will the Poles be in ten, in twenty years? When they can no longer be romantics and content with **reading Mickiewicz** [...]? (NKJP 22); [...] Włodzimierz, as befits a true Pole, rushed to greet his honoured guest. Although the **bread and salt***

¹⁷ If the use of the mother tongue rises to the rank of a classifying factor and gives rise to a feeling of superiority in the user, it may become a factor of exclusion (and at the same time discrediting a true Pole in the eyes of others), as signalled in the next part of the quotation: *Beware! it may turn out that you are then talking ungrammatical Polish and not Polish). i.e., better than, e.g., a Silesian (who wants his/her language to be recognised as a regional language [...]).*

¹⁸ One participant in the online discussion noted that this was rather a postulate, because: *Friend, You have missed some important characteristics of "a true Pole" [...]. It is not always easy to take care of the "cleanliness" of the language, too many words and phrases have become permanent in our everyday speech, e.g., TV, computer and even the Internet are not Polish words and nobody tries to convert them into Polish. (net. 3c); another was critical of other features: Hello, you wrote an ideal picture of a Pole and it is exactly the opposite, he/she knows history, which is usually hypocritical, depending on which side the wind blows, with this language it is even a pity to talk because I would have to point out every other word, national holidays are an opportunity for a riot and the flag is often a sheet scrawled with the names of villages of sports fan (net. 3d).*

in his parent's hands was replaced by a bottle of Absolute, the moment was solemn and touching. (NKJP 29); From the hands of "a true Pole", I received a leaflet from "Placówka" about the consequences of Poland's access into the European Union. [...] **"A shepherd boy behind a cow, hollyhocks by the roadside, roadside willows, children** going to school, a village in the distance", it brings a tear to the eye. (NKJP 31); Soon, true Poles spoke up, claiming that Valentine's Day is an ideologically foreign borrowing from the Americans. After all, we have our own Slavic festival of love – **the Kupała Night**. (NKJP 32); [...] there will be only "true" Poles left, **serfdom** will return, the **manors** will be rebuilt and everything will be as it was under the rule of the old, pure-blooded Poles. (NKJP 33); [...] and he/she is left with his/her companions praising **vodka, sour soup, Old Polish bigos and hussar charges**. (NKJP 34); The Parliamentary Club of the Non-Partisan Bloc in Support of Reform fully supports the draft resolution on the International Cooperative Day, advocating that our country should finally be governed by true Poles – **farmers raised in our native traditions**. (NKJP 35); And no true Pole **will swallow horse broth** or a piece of meat as he/she will "return" it and beat [...]. (NKJP 38); [...] sense of humour, **attachment to country and culture**, evoking good fun. (net. 4a); [A real Pole is one who...] **Knows** the history of his/her own country, its **culture**, and knows how to celebrate **Polish traditions** with dignity. (net. 4b); And I will say more, a real Pole [...] **knows the history** of his/her Nation (net. 3); [...] **on Easter Monday** you poured water [...], Christmas Eve dinner consists of **12 dishes** [...], there is no better party than a **Polish wedding** [...] (net. 8), "So if someone thinks it's worth being a Pole, they have to stand on the side that **defends traditional values** and wants to rebuild our reality to make it much fairer, although there has been a significant improvement here". [net. 12];

2.6. A Pole is considered only a native inhabitant of Poland (territorial criterion) or one who has Polish origins for many generations, mainly of nobility (social criterion): *First of all, he/she should have a proper family tree, i.e., be a Pole from his/her*

fathers. Personally, he/she does not have to be a grandfather. (NKJP 21); [...] There were voices on other sides demanding the so-called "indigenusness", trying to convince us that **only a "native" is a true Pole**, and a "non-native" is not really so, **as if Polishness were not a spiritual but a biological and animal category to be forced, as if I wanted to renounce**, for example, **those** Polish artists **whose families came here from other nations and became Poles**, such as that half-Orman Juliusz Słowacki and many, many others, without whose work I could not live. (NKJP 50); **I am the son of a prince** and a farm girl [...]. Why, then, do you not acknowledge me and my beloved, who is also a **noblewoman**, albeit only a homestead one, but with a coat of arms... After all, I am a true Pole, and not the son of a grocer, innkeeper, or usurer. (NKJP 45); Mr President [Wałęsa] assured everyone that, in contrast to his adversaries, he has been **a true Pole for many generations**, while people of Jewish origin were publicly called upon to reveal their identity. (NKJP 47); **"Wałęsa has Polish blood and has documents to prove it"**. "Where does anti-Semitism really come from? From where certain people are hiding, and it won't happen if there is honest, open play." **The most frequently cited motive for electing Wałęsa in the previous election** – 37 percent of respondents – was that he is **a "true Pole"**. (NKJP 48a); Yes, there are some evangelicals, Jews, Orthodox Christians and Muslims among Catholics, but let's be honest – they are not true Poles according to our right wing. **They are foreign bodies**, who are not allowed to have **ethnic Polishness**. [...] I am a non-believer. Of course, this is better than being a Jew or a Tatar, and much better than being a Roma, but still not enough to be a true Pole. (net. 15); We want to live in Poland, where Poles live, **feel in Polish, are Polish in flesh and blood**. (net. 19)

2.7. (in connection with 2.6.) has prejudices against representatives of other nations – aliens (bordering on xenophobia,¹⁹

¹⁹ In the sense of dislike or hostility towards foreigners and foreign culture – according to USJP (*xenophobia*).

identifying this alienness even by name), attributes to them bad intentions, various types of actions (intrigues, conspiracies) to the detriment of Poland and/or Poles:

– towards the Jews: [...] **does not like the Jews** – in a word, a 'true' Pole. (NKJP 2); [...] *after fifty years, he decided to visit his home town. And he reads the inscriptions on the walls in Polish: "The **Jews to the gas**" and in German "**Juden raus**". What can we say when **Hitler's articles on racial purity are today echoed by the term "true pure-blooded Pole"**? And when this title of pure and true Pole is denied even to those patriotic activists of Jewish origin who, alongside others, won Poland a return to freedom [...].* (NKJP 3d); "A True Pole" believes that the **Jews are in the lead in the editorial board**, who stole the magazine from the workers. He writes indictments on Mazowiecki and sends them around the union authorities. Where do such types blinded by fanaticism come from? (NKJP 18); And beautifully. A few months ago, **attacks on the 'Lithuanian' Miłosz, now on the 'Jew' Huelle**. [...] The translator of Polish literature, Renate Schmidgall, once told me about her meeting with the author of "Weiser Dawidek" as well as with Krzysztof Koehler and Maciej Niemiec: "I suddenly realised that I was talking to three Polish writers, the first of whom is called Huelle, the second Koehler, and the only true Pole in this group is ... German." (NKJP 40); In previous years, depending on the needs and what they preached, the most enlightened men of the Polish Church were called the **Jews on duty to beat**: Rev. Prof. Józef Tischner, Bishop Józef Życiński, Cardinal Franciszek Macharski, and Pope John Paul II. [...] There are not many Jews in Poland, but the **attitude towards the Jews is as far from indifferent as it was before the war**, when they were a minority of more than three million. Apparently the '**Jewish problem**' consisted not so much, and in every sense not only, in the presence of the Jews as in the **attitude towards the Jews**. (NKJP 4); "**You Israeli lackey! Mazowiecki can rule in the synagogue, not in Poland!** [...]" – I heard. There were also others. They would come up to me and explain in a fatherly manner that true Poles do not vote on '**such a one**' (here followed a gesture of stroking an imaginary beard).

"The Jews want to rule Poland! This whole government is **Jewish!**" (NKJP 3b); **the Jews steal** not only money but also time. As true Poles, we **ask our blood brothers not to let us snatch from our hands what we already possess.** There is an enemy hiding in the shadows, from behind the scenes affecting the fundamental right to birth and the sanctity of the sacrament of marriage between a man and a woman. (NKJP 46); The topic has not become outdated – according to a recent Demoskop survey, 36 percent of the public believes that **Jews have too much influence over political and economic life in Poland.** (NKJP 48b); KOR, Round Table are for them [the youth] completely indifferent terms, and yet it was our road to freedom. And here, on the one hand, Radio Maryja announces that they were not true Poles, but **some Jews** and Freemasons,²⁰ and on the other, young people are completely unmoved by that. (NKJP 19); [times before Solidarity] [...] **overt anti-Semitism** as in March 1968 [...]. In order to delegitimise the opposition as foreign to the nation, propaganda resorted to various tricks, for example quoting **Jewish names**, using various **periphrasis** such as "those who probably entered a church for the first time". Also, in frequent use was the phrase "true Poles", which party authors contrasted with oppositionists such as Adam Michnik and Seweryn Blumstein. (NKJP 42); [Bodo] "supported **the boycott of Jewish trade**". [...] he avoided contacts with the trade, as befits a True Pole [...]. (NKJP 20); Already in the second half of 1981, in the face of economic troubles and the increasingly intransigent attitude of the government, **anti-Semitic statements** began to appear, directed mainly against advisers to the movement who originated from the KOR, and in the "Mazovia" region, a group describing itself as "real Poles" undertook an open rebellion, accusing the union authorities of submissiveness to "cosmopolitans". (NKJP 51); Meanwhile, when someone said that he/she was a real Pole as opposed to "untrue Poles", that proved very useful. I was horrified for good when a large part of the Solidarity base accepted Marian

²⁰ According to a true Catholic Pole, Freemasons and people of other faiths constitute a (potential) threat to his/her sense of religiosity, his/her traditionally professed faith, his/her recognised values.

*Jurczyk's famous speech in Trzebiatów, in which he said that **Poland was not governed by Poles but by the Jews under changed names** [...]. A rather disgusting spectacle was taking place. **The accusation was that someone had a Jewish surname** [...]. (NKJP 49); – **How many Jews are there in Poland?** – I asked. – **Too many** – he answered. In his opinion, **the Jews predominate in the government, the Sejm and the Episcopate. President Wałęsa is a Jew who sympathises with the Germans.** True Poles, like Chairman Tejkowski, are subject to repression and political persecution. (NKJP 52b); And the aggressive "truthers" (true Poles), **demanding the blood of KOR and purging the union of the Jews?** (NKJP 27); But who is the **opposite of a true Pole**, this untruthful Pole? The answer is clear, although often not formulated directly. It is a Pole of Jewish origin or simply **a Jew**. [...] The communist authorities at one time deliberately tried to **replace a foreign-sounding surname with an indigenously Polish one**. (net. 13). In a narcissistic impulse, she added another, messianic amendment to this motto. Poland is the bride of God (Christ the King of Poland, the Virgin Mary the Queen of Poland), and the Poles – **not some insolent Jews** or other such – the chosen people. (net. 14); I went to see the film "Son of Shawl" for the second time this evening at the Muranów cinema in Warsaw. "A true Pole" was revealed in the auditorium. [...] This delightful gentleman gave a lecture this **evening on the Sonderkommando**, the ignorance of the film's director and his **despicable manipulations, and above all he exposed the conspiracy of the makers and distributors of "Son of Shawl" – a strike against Polishness. This film is anti-Polish! "A true Pole" knows what really happened in the Sonderkommando at Auschwitz-Birkenau.** Last December, at a special screening of Yael Hersonski's film about footage shot by the Germans in the Warsaw Ghetto in May 1942, "a true Pole", defending his alien civilisation by suggesting that the **Jews murdered themselves**, was revealed. [...] I think that "true Poles" are*

afflicted with a specifically Polish variety of AMD²¹ of the moral sense. [...] The **moral panic** into which "true Poles" fall erupts **whenever the subject of the Holocaust of the Jews appears on the public scene**. (net. 17); [...] the **Jewry, headed by Michnik and Blumstein**, dressed up the youth, pseudo-youth, in Auschwitz uniforms and confronted us, because it was the so-called *antifa*, i.e., a march against fascists. Sir, how can Poles who celebrate Independence Day be called fascists? And our authorities allow **the Jews** [...] to call us fascists. (net. 19);

– against other nations, especially Germans and Russians: **against the Saxons and against the Swedes, against Russia** too. In short, against the whole world. A true Pole. [NKJP 43]; Kazimierz Lewenko from Debrzno never regretted not flying into space. Who would believe that **Lewenko is a true Pole**? [...] Colleagues from the regiment even advised him **to change his surname to one ending in "ski"**, but he decided not to do so. (NKJP 44); How can one distinguish "a true Pole"? – First of all, **by his worries**. This is an individual who is constantly and painfully experiencing the terrible problem that not everyone has yet managed to get covered in mud. So much work: receipts, files, studying CVs, looking into trousers, pouring out mud... **And here there is always a not spit "European" peeping out from behind the corner**. [...] For him [Jan Węglowski from "Nasz Polska"], **Miłosz** is not an inhabitant of the Małopolska region, because he **is not Polish** at all. "He does not feel absolutely Polish, but Lithuanian, in "Tygodnik Pow-szechny" in 1996 he officially confessed: "I would have nothing against being called Lithuanian". [...]" (NKJP 48); In the final phase of the [election] campaign the spin doctors gave the face of the Grand Master of the Teutonic Knights the **features of a grandfather from the Wehrmacht**. In this way Donald Tusk was perfidiously marked with the stigma of the Stranger. (NKJP 56); According to Kaczyński, Morawiecki and company, a Pole is only someone who supports their policies and conservative, Catholic ideology. **Whoever does**

²¹ AMD of the moral sense is a metaphor because AMD (from the English name: Age-related Macular Degeneration) leads to damage to the retina and, in consequence, even loss of sight.

not support, is not a Pole, and in a stronger version is a German. [...] From this hysterical alternative: either the West or Polishness (as if we were going back three hundred years to the Saxon times and the dispute over the kontush versus the tail-coat) [...] A **statesman** is a person who **opposes the Germans.** (net. 12); to help, they invoke the noblest patriotic traditions and slogans, e.g. (this about Angela Merkel and Martin Schulz) "**no German will spit in our face**". They consider it a test of patriotism and honour to "**de-Sovietise**" and "**de-Russify**" in such a way as to neglect the cemeteries of Soviet soldiers, and in Sandomierz, Kraków or Częstochowa (no Bolshevik will be our patron saint) to remove any trace of the Red Army commanders who saved these cities with their tactical decisions. [net. 14];

– towards strangers (in general), Others, often dividing (and/or antagonising) the Poles into "us" – "them": **Selling land to strangers is a crime to me.** I want to know if the starost feels the same way? Once again, the councillor accused the starost of not acting like a true Pole. (NKJP 41); The letter was commented on by Jarosław Kaczyński, editor-in-chief of "Tygodnik" and a leading politician of Wałęsa's camp [...]. In the "Communication of the Plenipotentiary of the Centre Agreement for Pińczów and the Region" ("Gazeta", 13 XI 1990) we read: "We, true Poles, seeing what is happening in the country, must already today strive to speed up the elections. Today, we already know **who is ours** and who is not. **Power has to get into the hands of true Poles** – Christians, therefore the electoral fight will be hard and relentless – to the death". (NKJP 3c); So, who will the Poles be in ten, twenty years? When they can no longer [...] **turn away from strangers, especially the Jews.** (NKJP 22); One of them explained to me that, above all, one **should oppose selling Poland to foreigners and the settlement of the Germans and the Jews in our country.** (NKJP 52a); And yet, that is where our roots lay in the nineteenth century, where a consistent distrust of others arose, **a desire for the nation to be 'pure'** and for all true Poles to be alike. [...] **hatred of others and rejection of them** [...]. [...] everyone was to resemble one another – [...] act as members of one family. **Lack of respect for otherness** – it is

what characterises a Pole even today, and was sown then, or perhaps even earlier... (NKJP 22); I am convinced that the "Mazurek Dąbrowskiego" will survive. And even true Poles will not trace anything here, because the author of the music of our anthem is unknown. So, it is difficult to make the accusation that he was **a Freemason, a Jew, a bigamist**, or something else. (NKJP 53); We have become extremely distrustful. **We do not want to support refugees, we are afraid of otherness**. However, we are one of the few countries where there is the so-called cultural homogeneity. (net. 2); **Normal** is what is **traditionally Polish**. What is **abnormal** is what is **in the West** (net. 12); The historical mission of Poland and the Poles is: to be the Bulwark of Christianity. The **Poles' superiority** – especially moral and "evangelical" superiority – over the whole surroundings, and in Poland itself, on their own territory, **over all outsiders and strangers** – those who have settled here for generations, for centuries, and those who push their way towards us, sacrilegiously, blasphemously polluting our religious-patriotic virginity. (net. 14); [...] I have no affinity with the right wing because of the uncivil behaviour displayed by many 'true Poles' wearing black T-shirts with pictures of crowned eagles and inscriptions reading 'cursed soldiers'. I am particularly repelled by right-wing racism and anti-Semitism, i.e., **xenophobia** in the broadest sense. (net. 15); A drunken man verbally attacked **two Asian women**. Their **background did not please** the self-proclaimed "true Pole", who told them to "get the hell out of our country". [...] **Poland is for the Poles** and that they were definitely illegally here. The story was described on Facebook by Maria Cywińska, author of the blog Lumpiata. A tattooed, drunk man in a black T-shirt. No hair. The exact same guy accosted me, my boyfriend and his dad, saying that we were supposedly talking about **the Balkans** [...] he started following us and shouting after my boyfriend if he was a real Pole. He came close **with his fists clenched and told him to say the words of the Polish anthem**. That went on for a while, at the Marymont Metro stop we ran away and he was shaken by the door. (net. 16);

- sees external (coming from abroad) cultural and political threats: [*From the hands of "a true Pole" I received a leaflet of "Placówka" about the consequences of Poland's entry into the European Union.*] *And what will it be like? **Joining the Union is a mortal threat** to the Nation, the economy, "the Spirit and, of course, the roots."* (NKJP 31); [...] *the true Poles will treat the day of 13 December 2002 much worse. Patriots from the League of Polish Families consider the end of negotiations in Copenhagen, **opening the door to the European Union for Poland, as an act of national treason, selling the homeland to the Freemasons, and Targowica.*** (NKJP 69);

2.8. Reconstructing the linguistic picture of a *true Pole*, it is furthermore possible to notice certain features represented by them, which are considered:

2.8.1. as advantages:

- hospitality: *the Poles are **hospitable** and cheerful, unfortunately there are snobs.* (net. 4f); **Hospitality**, patriotism, religiousness (net. 4g); *You can learn diligence, loyalty, **hospitality** from us [...]* (net. 2);
- chivalry towards women: [...] *Major Morawski changed his tone, to "almost lyrical", adopting the attitude of a true Pole. He said that he felt **great pity for the dear ladies**', that he was also a Pole and understood their ideological motives [...].* (NKJP 37);
- the family spirit:²² [*Family Affairs Minister Marian Kapera, a patriot and a true Pole, wanted in this way [by subsidising Viagra] to **increase the birth rate of the nation**, because our fertility rate is decreasing [...].* (NKJP 36); *he was never a father, at most a sperm donor, he begotten and then immediately went to the reserves, he did not take me to the seaside or to the mountains, he did not show me Częstochowa or the Częstochowa Jurassic Highlands, he did not read books to me, although it was*

²² In the meaning of being familial: loving family, enjoying family life, caring for its development – cf. UJSP *family spirit* m. 2.

*his bounden duty, he did not teach me anything, he only gave me money, instead of catechism he brought sausages, treasures from the shop behind yellow curtains, but he never prayed for us, he **was not a true Polish father**. (NKJP 39); **Our greatest value is the family**. (net. 2); [...] **the family is an arch-Polish value** (net. 12); A typical right-wing family consists of a man – **the head of the family**, his wife – the **housewife**, and **a bunch of children**, preferably equally boys and girls. (net. 15);*

2.8.2. as disadvantages:

– predilection for coarse²³ jokes: [*For with the new western world there also came the new*] [...] **coarse jokes** were replaced by thin jokes about blondes [...]. (NKJP 30);

– loud, inelegant, uncultured behaviour: [...] *it is difficult to deduce anything about the appearance of a true Pole, except that he/she speaks **loudly**, often accuses, and is always right*. (NKJP 22); *All the invited guests [...], as true Poles do, laugh, "**laugh like a drain** I think". [...]* (NKJP 34); *In a crowded restaurant there was a pair of cave people with a **few-year-old savage**, who, **shouting**, dashed between the tables, **pushed and shoved** other children, and banged on all the equipment and windows with a stick in his hand. The cave-dad **with a mouth full of food was burping**, grunting, and snorting every now and then – encouraging his son to fool around. The mousy mummy only nodded her empty head with admiration and approval. At that time, in front of the restaurant, **a well-dressed middle-aged buffoon with a large dog**, obviously without a leash or muzzle, was celebrating his walk [...]. Its owner, as befits a true Pole, did not deign to clean up the poop. (cf. NKJP 62);*

– conviction of one's own right: "*True Poles*" are always – in their view – **perfectly prepared**. It's not even that they have their own books and historical fact sheets (dates, names, events and "trick questions"). **They just know, unlike everyone else who is wrong**. (net. 17);

²³ See coarse in USJP m. 5: *obscene, gross, vulgar wit, joke*.

- having complexes (consequently expecting approval from others): *This is the result of the private chimeras and resentments of a few people who are afraid of the world, **have complexes**, don't know languages. So, they tell the voters that a true Pole is out for himself/herself. (NKJP 60); Mr President Lech Wałęsa is the only that can save us, the Poles. [...] A true Pole, [...] a **frequent politician in the world**, an honest man, with Mr President **they count as with no one from among the Poles**, naturally excluding only the Holy Father, all over the world. (NKJP 5); [About Chopin] "A real" Pole [...] **a European without any complexes**, making friends with the greatest people of his era. (NKJP 65);*
- lack of knowledge of foreign languages²⁴ (perhaps partly due to a peculiar concept of patriotism and love for the mother tongue, and aversion to Foreigners): *This is the result of the private chimeras and resentments of a few people who are afraid of the world, have complexes, and **do not know languages**. [...] (NKJP 60); [About Chopin] A "real" Pole until the end **speaking not the best French** [...], (NKJP 65); The message that Kaczyński is a true Pole, was reinforced by the myth of the Warsaw Uprising and statements: I do not go abroad, I do not talk to Germans, **I do not know foreign languages**. (NKJP 56);*
- propensity to:
- *drunkenness: *[...] a true Pole – **drunk** and anxious about the homeland. (NKJP 22); Does the name Piotr Cyrwus mean anything to you? [...] Suffice to say – Rysiek from "Klan", and every true Pole knows who he means. [...] However, the biggest drawback to his role as Rysiek from "Klan" is the **fact that almost everyone wants to have a drink with him**. (NKJP 59); He has good blood. Bright hair. No suspicion on my part. Only he **does***

²⁴ This allegation is increasingly overcome by contemporary opinions: The pharmacist ponders for a while, asks in **fluent English** if this is for adults. He instructs how to take the medicine and writes all this on a Polish-language leaflet – in English. One euro for a true Pole (NKJP 66); "One can certainly learn from Poles the motivation to learn foreign languages". [...] Moreover, as indicated by the ranking "English Proficiency Index 2013", **the Poles are among the ten best English-speaking nations in the world.** – points out Anna Białous-Griffiths, a psychologist and personal development coach in the UK. (net. 2), which, however, by this type of emphasis, confirms that the trait is considered frequent/constant for a true Pole.

not drink like a true Pole [...] (NKJP 46); [...] *everyone is obliged to **get drunk from time to time, otherwise he is not a true Pole.*** (NKJP 64); **Strong head for drinking** :P (net. 4); *And finally, a special candidate: he does strange business, goes bankrupt and comes out on top, **drinks**, swindles [...] – in a word, a "true" Pole.* (NKJP 2);

*criticism: [...] *there is **nothing that a true Pole cannot complain about.*** Such a manifestation of our **distrust** of everything. (NKJP 61);

*complaining: *A true Pole **likes to complain**:-)). Although there is often no reason for it:-))* (net. 3); [...] *our vices, such as: xenophobia, combining, **complaining**, excessive directness, even rudeness. On top of that there is our eternal pessimism and criticism of others* (net. 2);²⁵ [...] **complaining** is a way to vent emotions, to tighten ties (net.8);

*aggression: *I won't respect someone who hurts me because **I've already been ripped off once** by one who called himself "**a true Pole**": and added "that I wouldn't mess with skins".* (NKJP 63a); [...] **he will barge into her house wearing a balaclava?** only bald "true" Poles are capable of such things (NKJP 63b);

*philistinism: *One mustn't flatter **a babbitt**,²⁶ who considers himself/herself a true Pole.* (NKJP 67);

*negligence: *The face of a true Pole always shows a pleasant, cheerful smile. It expresses the **peace of mind** that comes from a sense of **well unfulfilled duty**...* (NKJP 68);

*intolerance: **With time, a person** notices that the world is not black and white. He/she abandons youthful fanaticism and **learns to tolerate others.** Unfortunately, not everyone gets over it and we have the dubious pleasure of sometimes hearing about Pajak, Bubel or other true Poles. (NKJP 63); **Intolerance** was accompanied, as I said, by chauvinism. [...] (NKJP 57); *And seriously, it may be a stereotype, but I associate a Pole with a person who is [...] a bit **intolerant**, it is difficult for him/her to tolerate /*

²⁵ In this quotation, the author indicates many other characteristics, so it is only conventionally placed in this group.

²⁶ See a *babbitt* in USJP: c) contemptuously a narrow-minded person, backward, dark, limited; obscurantist.

accept "otherness", vindictive :) (net. 4); I also have bad associations with "a true Pole" and, in my opinion, it is not a nice portrait of THAT Pole. And it is a complete paradox that **a true Pole claims to be tolerant, but it has nothing to do with reality.** (net. 18); [CONTEXT: "[...] And I thank God for the fact that this rainbow in the sky cannot be burned, destroyed or used to stigmatize and humiliate others," **Stanisław Soyka** wrote on Facebook on August 12, posing with a rainbow flag. [...] Three days later, Soyka was one of the participants of the show "**1920. Wdzięczni Bohaterom**" (Indebted to Heroes), during which he performed "Tango Warszawo" (Tango Warsaw). However, the artist's participation in the TVP event did not please many right-wing Twitter users:] What do I see **rainbow Soyka reached out for money to TVP, recently he spit on the intolerant Polish government, but hypocrisy, money from public TV does not stink.** (net. 20a); The worst thing is that first **they spit on Poland and the Poles, and then they participate in celebrations important for the country** and still get paid for it. (net. 20b); When Soyka came out yesterday, I switched the TV channel. **His and Wyszconi's participation in this show is, in my opinion, an insult to my ancestors and other Poles fighting in this war** (to say it mildly). (net. 20c);

*envy: A true Pole is an individual who **has envy written in his genes and passes it on from generation to generation.** It is not since time immemorial that we have a proverb about a dog in the manger. (net. 3);

*egoism: [...] **the primitive egoism** of the modern Poles – Property is for the right wing a sanctity bordering on idolatry. The short phrase "it's mine, so I won't give it away!" defines the attitude of every true Pole to such social values as solidarity, for example. (net. 15);

*others: **He/she hates** everyone who has a better life and does not believe in God. This is such a Polish trait. [...] **Battle-worn, eternally unhappy, grouchy.** (net. 4); For my part, I would add **Catholic hypocrisy, envy, thievery and cunning,** unfortunately [...] (net. 3); **the genes of a thief** (because in the past it was a patriot's duty to steal from the invader), **fraud and a ca-**

reer at any price (fighting for power, which ensured a prosperous life), rather than personal culture and honesty (net. 9); And in God-fearing everyday practice it means something else: to **exult in honour whenever someone brings errors to our attention**, whenever they expect or demand from us what is inconvenient for us. [...] And the ordinary "true Poles" have a similar sense of honour. They feel no dishonour when they **litter** their native forest, when they **burn trash** and polysterene in their chimneys, when they **do not clean up dog poop** in front of their own gate. (net. 14);

2.9. Characteristics of a Pole/Polish people that are nowadays most often perceived (as positive) by other nations, e.g., **fighting spirit, sense of independence, creativity**; [...] We usually **look for effective, inexpensive and at the same time innovative solutions**. In Great Britain, over 14 thousand Polish companies are established every year. Polish IT specialists, for instance, are among the best in the world. [...] Paweł Lajszczak of Financial Brokers said: "In general, the Poles are **very determined in specific tasks** that are given to them, they perform them meticulously and ingeniously". [...] Not only are we **eager to learn languages, but also any theoretical and practical knowledge**. [...] They usually do not specialise in one field. They are able to solve problems from different areas independently. Slogan: a Pole can. [...] A Pole is not afraid of any work – such a belief about our nation prevails in almost every country in the world. [net. 2].

Interestingly, among the above-mentioned characteristics (from 2.9.), there is no reference to nationality, origin (territorial and social), and religion. It proves that for other nations, they do not matter too much, and what matters is the potential (pragmatic, intellectual and social) inherent in a person.

3. Conclusion

From the review of about 100 texts, a polarised image emerges, which on the one hand shows that *a true Pole* is:

- a Catholic (defender of the Church and Christian values, cooperating with the hierarchy of the Church, granting great trust to the Catholic media; Lech Wałęsa, Tadeusz Rydzyk, Edward Frankowski, St. John Paul II, Henryk Jankowski, for example, were recognised as models of a Catholic understood in such a way and as authorities);
- feeling strong ties with his/her nation (a nationalist);
- a patriot (ready to sacrifice for the homeland, proud of the historical role of Poland, attributing special merits to the Poles in saving the world);
- accepting the Polish language as native;
- valuing Polish culture, cultivating native customs;
- considering only an indigenous inhabitant of Poland as a Pole (territorial criterion) or one who has had Polish origins for many generations, mainly of nobility (social criterion);
- hospitable;
- chivalrous towards women;

but also:

- an opponent of evolutionism (due to his/her religious beliefs and the idea of creationism);
- a nationalist (as an extreme of a nationalist);
- convinced of the existence of enemies of Poland;
- an opponent of socialism/communism, Nazism, liberalism;
- a person with prejudice against the representatives of other nations (called Strangers, and identifies the alienness even only by name), attributing to them bad intentions, various types of actions (intrigues, conspiracies) to the detriment of Poland and/or the Poles, especially the Jews, but also the Germans, the Russians, immigrants, people of other beliefs, religions and orientations (Others), which gives him/her grounds to divide and/or antagonise the Poles into "us" – "them";
- looking for external (coming from abroad) cultural and political threats;

- with a predilection for coarse jokes;
- behaving loudly, in an inelegant and uncultured manner; convinced of his/her own right;
- with complexes (especially when compared to other nations – representatives of the West);
- without knowledge of foreign languages;
- prone to drunkenness, criticism, complaining, aggression (especially in the defence of his/her rights and values), philistinism, negligence, intolerance, envy, egoism, jealousy, hypocrisy, feeling offended, etc.

That set of mostly negative characteristics should be complemented by elements of the contemporary hetero-stereotype (of 2018): independent, creative, inquisitive, self-reliant, hard-working (see 2.9).

The characteristics indicated in Jakub Żulczyk's article correlate with the above-mentioned juxtaposition:

The first view holds that a true Pole is a **strong, well-built, brave** grandson of insurgents. His heart is a great vessel, filled to the brim **with pride, honour, nobility, and love for the homeland**. A true **Pole buys only Polish products and without hesitation would give his life in a war with the Russians**, and for the sake of the cause he could even provoke such a war himself. By the side of a True Pole there is a True Polish Woman, ready to give birth to another True Pole at any moment, stroking his tired head and reading "**Pan Wołodyjowski**" **to sleep** after a day-long battle with the Russians, the Third Republic, the Smolensk liars, and all the rest of more or less hidden **enemies of Poland**.

The second view of a true Pole is just the opposite. According to this diagnosis, a true Pole is **middle-aged, has a big belly, a sweaty shirt, a moustache, hates everything and especially the Jews**, spends his afternoons greasing himself with **pork knuckles** while **watching 'Holiday Diaries'**, **beating children** with a belt, and **spitting** at the cathode ray tube when Donald Tusk appears on the screen. It should be added that a True Pole **steals cars, listens to disco polo** even when asleep, and **a priest** is the most important figure in his life. (net. 10)

Interestingly, some features from the bundle abstracted from the selected texts are also found in the description of the stereotype of the NATION (Żywicka 2015: 133-146). Some similarities emerge from the analysis of the ASA survey presented by the author:

- ideological/worldview – both concepts under study occupy important places in Polish culture, stemming from a sense of community;
- conceptual and cognitive one: both concepts show a predilection for (re)interpretation, or rather susceptibility to it, which was noticed by Jadwiga Puzynina, who wrote about *a nation* as an expression "particularly dependent on the political and social views of the language users, and at the same time always conceptually vague" (1997: 383).

Both notions can be considered from **different perspectives**, e.g., cultural (including religious), social (state-citizen), ideological, national (based on the so-called ethnic purity), which imply the extraction of alternative *differentia specifica* related to, among others, origin, territory, language, language, and the language of the nation. They imply the extraction of alternative *differentia specifica* related to, e.g. origin, territory, language, customs, perception of history/national self-knowledge, and sovereignty [cf. Puzynina 1998: 263] and which are not subject to multi-generational petrification, but, on the contrary, their configurations or modifications are influenced by **subjective and/or external** causal factors (e.g., migration, greater acceptance of bivalence) (cf. Bednarczuk 2009: 85).

In 1990, young people identified a "true" nation above all with its **language** and **common history**. Next, the respondents mentioned: cultural achievements of the nation, cultivation of traditions and customs, sense of patriotism, national consciousness (national bonds); they also exposed social features such as community (language, culture, goals), unity, solidarity, sovereignty (cf. Żywicka 2015: 135-137).

In 2000, the sense of **community**, which is created by society and having one's **own history**, was considered to be the most crucial in the picture of a "true" nation; ideological features, e.g., patriotism, a sense of national identity, and attachment to symbols, were more frequently emphasised, the essence of a "true" nation was associated with religion, cooperation, caring for the well-being of others, for one's own language and traditions, unity, solidarity, adding the territorial factor – areas inhabited by a given group of people (Żywicka 2015: 137-138).

In 2010, the dominant features focused around **the sense of unity and patriotism**; the value of national symbols, the important role of language, cultural heritage, and the need to cultivate traditions, customs were indicated; the impact of history on national awareness was emphasised; cooperation, solidarity, harmony, mutual help, and the influence of commonly held values and religions were appreciated (Żywicka 2015: 138-140).

The linguistic picture of a true Pole reconstructed in this way becomes the basis for further interpretations. It seems that each (new) facet enables another re-interpretation, which, however, requires additional research that will be presented in another study (this volume).

References

- Bańko, Mirosław (ed.) (2000). *Inny słownik języka polskiego* (ISJP) Vols. 1-2, Warszawa: PWN.
- Bartmiński, Jerzy (ed.) (2006). *Język – wartości – polityka. Zmiany rozumienia nazw wartości w okresie transformacji ustrojowej w Polsce. Raport z badań empirycznych*. Lublin: Wydawnictwo UMCS.
- Bartmiński, Jerzy, Jolanta Panasiuk (1993). "Stereotypy językowe". In: Jerzy Bartmiński (ed.). *Encyklopedia kultury polskiej XX wieku*. Vol. 2. *Współczesny język polski*. Wrocław: Wiedza o Kulturze, 363-388.
- Bednarczuk, Monika. (2009). "Rozumienie narodu i jego profilowanie we współczesnym języku polskim". *Etnolingwistyka* 21: 85-101.
- Dubisz, Stanisław (ed.). (2005). *Uniwersalny słownik języka polskiego* (USJP). Warszawa: PWN.

- Dunaj, Bogusław (ed.). (2001). *Słownik współczesnego języka polskiego* (SWJP). Vols. 1-2. Warszawa: Reader's Digest.
- Markowski, Andrzej (ed.). (1995). *Praktyczny słownik poprawnej polszczyzny nie tylko dla młodzieży* (PSPP). Warszawa: PWN.
- Mazurkiewicz, Małgorzata (1991). "Stan prac nad słownikiem aksjologicznym". In: Jadwiga Puzynina, Jerzy Bartmiński (eds.). *Język a Kultura*. Vol 2. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 257-264.
- Opaliński, Edward (1995). *Kultura polityczna szlachty polskiej w latach 1587–1652: system parlamentarny a społeczeństwo obywatelskie*. Warszawa: Wydawnictwo Sejmowe.
- Puzynina, Jadwiga (1997). *Słowo – wartość – kultura*. Lublin: Wydawnictwo KUL.
- Puzynina, Jadwiga (1998). "Struktura semantyczna narodu a profilowanie" In: Jerzy Bartmiński, Ryszard Tokarski (eds.). *Profilowanie w języku i w tekście*. Lublin: Wydawnictwo UMCS, 259–276.
- Rulka, Janusz (2001). *Wychowanie obywatelskie*. Warszawa: PWN.
- Sobol, Elżbieta (ed.) (2002). *Nowy słownik języka polskiego* (NSJP). Warszawa: PWN.
- Sobol, Elżbieta (ed.) (2001). *Popularny słownik języka polskiego* (PSJP). Warszawa: PWN.
- Szymczak, Mieczysław (ed.) (1999). *Słownik języka polskiego*. Vols. 1-3. Warszawa: PWN.
- Żywicka, Beata. (2015). "Obraz NARODU w świadomości językowej młodego pokolenia Polaków po roku 1989". *SŁOWO. Studia językoznawcze* 6: 133-146.

Netography

- Preambuła – Konstytucja Rzeczypospolitej Polskiej [Konst.] / Preamble – Constitution of the Republic of Poland [Const.] (arslege.pl) [accessed: 10.06.2021]
- net. 1²⁷ = <https://www.tygodnikprzeglad.pl/prawdziwy-polak-proba-definicji/> Witold Starzewski, *Prawdziwy Polak (próbą definicji) / A true Pole (attempt at definition)*, 11.11. 2019 [accessed: 20.07.2020]
- net. 2 = <http://www.damosfera.com/7-cech-narodowych-Polakow-czyli-jacy-jestesmy....html>, *7 cech narodowych Polaków / 7 national characteristics of the Poles*, 12.12.2018 [accessed: 20.07.2020]

²⁷ It corresponds to designations in the text, e.g., as net.1, net. 2, etc.

- net. 3 = <https://forum.interia.pl/kto-to-jest-prawdziwy-polak--tematy,dId,2645205> *Kto to jest prawdziwy Polak?* – Interia – forum dyskusyjne / *Who is a true Pole?* – Interia – the discussion forum, 20.05.2015 [accessed: 9.09.2020]
- net. 4 = https://zapytaj.onet.pl/Category/002,003/2,12033928,Jak_ie_cechy_ma_wedlug_was_prawdziwy_Polak.html, *Jakie cechy ma prawdziwy Polak?* – wypowiedzi internautów / *What qualities does a true Pole have?* – Internet users' statements, 10.06.201 [accessed: 9.09.2020]
- net. 5 = https://www.wsjp.pl/index.php?id_hasla=722&ind=0&w_szukaj=Polak+ *Wielki słownik języka polskiego / Great Polish Language Dictionary*, P. Źmigrodzki (ed.) (WSJP) [accessed: 12.09.2020]
- net 6. = https://sjp.pl/Polak_Slownik_jezyka_polskiego (SJP) / *Polish Language Dictionary* (SJP)
- net. 7 = https://crispa.uw.edu.pl/object/files/5890/display/JPEG_Slownik_jezyka_polskiego/ *Polish Language Dictionary*, J. Karłowicz, A. Kryński, W. Niedźwiecki (eds) (SW), vol. 4, Warsaw 1908, p. 539 [accessed: 16.09.2020]
- net. 8 = <https://www.polityka.pl/galerie/1714300,1,18-cech-charakterystycznych-tylko-dla-polakow.read> 18 cech charakterystycznych tylko dla Polaków / 18 features characteristic only for the Poles [accessed: 16.09.2020]
- net. 9 = <https://mazgula.pl/aktualnosci/wybory-ujawnily-wstydliwogeny-polakow/> Adam Mazguła, *Komitet Wyborczy Wyborców / Electoral Committee*, 16.07.2020 [accessed: 16.09.2020]
- net. 10 = <https://facet.onet.pl/warto-wiedziec/po-czym-rozpoznać-prawdziwego-polaka/t7lvqn1> Jakub Żulczyk, *Po czym rozpoznać prawdziwego Polaka?* / *How to recognise a true Pole?* – Onet 25.04.2013 [accessed: 16.09.2020]
- net. 12 = <https://oko.press/jeden-narod-jeden-kosciol-jeden-kaczynski/> Piotr Pacewicz, *Jeden Naród, jeden Kościół, jeden Kaczyński / One Nation, one Church, one Kaczyński*, *Oko.press* 21.07.2020 [accessed: 16.09.2020]
- net. 13 = <https://ksiazki.wp.pl/szczucie-zydem-i-moralnosc-kalego-6145577583786113g/5> Mateusz Witkowski, *Szczucie Żydem i moralność Kalego*, portal WP.pl – nawiązanie m.in. do wydania *Złej mowy* M. Głowińskiego / *Joking with a Jew and Kali's morality*, portal WP.pl – reference, among others, to the issue of *Zła mowa / Bad speech* by M. Głowiński, 16.11.2016 [accessed: 16.09.2020]

- net. 14 = <https://trybuna.info/opinie/bog-honor-dziczynna/> [no author], *Bóg, honor, dziczynna / God, honour, game*, Dziennik Trybuna 21.06.2017 [accessed: 16.09.2020]
- net. 15 = <https://trybuna.info/opinie/nie-prawica-nie-lewica-tylko-indywidualizm/> [no author], *Nie prawica, nie lewica, tylko indywidualizm?* / *Not the right, not the left, but individualism?* Dziennik Trybuna 28.06.2016 [accessed: 16.09.2020]
- net. 16 = <https://www.o2.pl/artykul/prawdziwy-polak-zatakowal-kobiety-warszawskim-metrze-6035833229742721a> Jan Muller, *„Prawdziwy Polak” zaatakował kobiety w warszawskim metrze / "A true Pole" attacked women in the Warsaw metro*, portal o2.pl 11.09.2016 [accessed: 16.09.2020]
- net. 17 = <https://hiperrealizm.blogspot.com/2016/01/Jacek-Leociak-Przed-czym-prawdziwi-Polacy-chca-sie-bronic/> Jacek Leociak, *Przed czym „prawdziwi Polacy” chcą się bronić? / What do "true Poles" want to defend themselves against?* blogspot Jewish.pl, 31.01.2016 [accessed: 16.09.2020]
- net. 18 = <http://joanna-kaffe-latte.blogspot.com/2015/10/prawdziwi-polacy.html> blogspot Joanna – *Świat według mnie / Joanna – The world according to me*, 17.10.2015 [accessed: 16.09.2020]
- net. 19 = <https://www.mojanorwegia.pl/video/8437/prawdziwy-polak-na-marszu-niepodleglosci.html> Moja Norwegia. Portal Polaków w Norwegii, *Prawdziwy Polak na Marszu Niepodległości / My Norway. Portal of the Poles in Norway, A true Pole at the Independence March*, 18.11.2015 [accessed: 16.09.2020]
- net. 20 = https://muzyka.interia.pl/wiadomosci/news-skandaliczne-wpisy-pod-adresem-stanislawo-wojtyly-pozszlo-o-tec,nId,4673512#utm_source=paste&utm_medium=paste&utm_campaign=chrome [no author], *Skandaliczne wpisy pod adresem Stanisława Soyki. Poszło o tęczałą flagę / Scandalous posts against Stanisław Soyka. It was about the rainbow flag*, Interia.pl 16.08.2020 [accessed: 16.09.2020]
- NKJP – *Narodowy Korpus Języka Polskiego / National Corpus of The Polish Language* http://nkjp.uni.lodz.pl/ParagraphMetadata?pid=32dc5f02515f8f99577b6dbf90c7be3c&match_start=423&match_end=438&wynik=36#the_match [accessed: 10.07.2020]
- NKJP 1 = *Kultura* no. 5/500 1989
- NKJP 2 = *Gazeta Wyborcza* 2. 11.1995
- NKJP 3 = *Gazeta Wyborcza*, 26.08.1995
- NKJP 4 = Jerzy Sławomir Mac, *Wprost* no. 46, 1997
- NKJP 5 = *Onet.pl – Rozmowy / Talks* 22.09.2003

- NKJP 6 = Gazeta Wyborcza 26.06.1993
- NKJP 7 = Polityka no. 2702, 2009
- NKJP 8 = Polityka no. 2290, 2001
- NKJP 9 = Marek Harny, Zdrajca, 2007
- NKJP 10 = Usenet – pl.soc.religia, 1997
- NKJP 11 = Adam Szostkiewicz, *Jezus by się dziwił, co wyprawiamy / Jesus would be surprised what we are doing*, Polityka no. 2496, 2005
- NKJP 12 = Polityka no. 2608, 2007
- NKJP 13 = Gazeta Wyborcza, 26.09.1997
- NKJP 14 = Słowo Polskie Gazeta Wrocławska, 21. 10. 2004
- NKJP 15 = Polityka no. 2485, 2005
- NKJP 16 = Gazeta Wyborcza, 10.07.1995
- NKJP 17 = Elwira Kucharska (2008). *Siła przyciągania, czyli polskie losy / The power of attraction, or Polish fates*, Lublin: Wyd. Norbertinum
- NKJP 18 = Waldemar Kuczyński (1981). *Burza nad Wisłą: dziennik 1980-1981 / Storm over the Vistula river: diary 1980-1981*; Warsaw: Wyd. ISKRY
- NKJP 19 = Gazeta Wyborcza, 9.01.1999
- NKJP 20 = Polityka no. 2228, 2000
- NKJP 21 = Dziennik Polski, 25.11.2000
- NKJP 22 = Bronisław Świdorski (2007). *Asystent śmierci. Powieść o karykaturach Mahometa, o miłości i nienawiści w Europie / Death assistant. A novel about caricatures of Muhammad, about love and hate in Europe*, Warsaw: Wyd. W.A.B.
- NKJP 23 = Dziennik Zachodni, 24.04.2010
- NKJP 24 = Stanisław Berenda-Czajkowski (2001). *Dni grozy i łez / Days of horror and tears*, Warsaw: Wyd. Czytelnik
- NKJP 25 = Wieści Podwarszawskie no 51, 2008
- NKJP 26 = Torańska Teresa (1985). *Oni / They*, Warsaw: Wyd. Świat Książki
- NKJP 27 = Mariusz Wilk (2007). *Wilczy notes / Wolf's notebook*, Warsaw: Wyd. Noir sur Blanc
- NKJP 28 = www.forumowisko.pl, 17.04.2008
- NKJP 29 = Dagmara Póltorak (2010). *Tak trudno być mną! / It's so hard to be me!*, Warsaw: Wyd. W.A.B.
- NKJP 30 = Józef Łoziński (2006). *Holdng i reszta albo: Jak zostać bogatym w biednym państwie / Holding and the rest or: How to become rich in a poor country*, Warsaw: Jacek Santorski & Co Agencja Wydawnicza

- NKJP 31 = *Wieści Podwarszawskie* no. 12, 2003
- NKJP 32 = *Polityka* no. 2692, 2009
- NKJP 33 = www.forumowisko.pl, 9.03.2006
- NKJP 34 = *Polityka* no. 2627, 2007
- NKJP 35 = Sprawozdanie stenograficzne z obrad Sejmu RP z dnia 20.04.1995 / Shorthand report on the session of the Sejm of the Republic of Poland on 20 April 1995
- NKJP 36 = Tomasz Jastrun (2003). *Gra wstępna rozpisana na 62 felietony / Foreplay written for 62 columns*, Warsaw: Jacek Santorski & Co Agencja Wydawnicza
- NKJP 37 = *Gazeta Ubezpieczeniowa*, 2003
- NKJP 38 = Ryszard Sługocki (2008). *Na przekór i na bakier / In spite of and at odds*, Warsaw: Wyd. Branta
- NKJP 39 = Marek Kochan (2007). *Plac zabaw / Playground*, Warsaw: Wyd. W.A.B.
- NKJP 40 = Wojciech Wencel, *Ozon* no. 23, 2005
- NKJP 41 = *Wieści Podwarszawskie* no. 49, 2004
- NKJP 42 = *Polityka* no. 2237, 2000
- NKJP 43 = *Polityka* no. 2484, 2004
- NKJP 44 = *Dziennik Polski*, 2005
- NKJP 45 = Andrzej Zaniewski (1997). *Król Tanga / King of Tango*, Warsaw: Wyd. TRIO
- NKJP 46 = Mirosław Nahacz (2005), *Bocian i Lola / Stork and Lola*, Sękowa: Wyd. CZARNE
- NKJP 47 = *Gazeta Wyborcza*, 4.02.1993
- NKJP 48 = *Polityka* no. 2254, 2000
- NKJP 49 = Adam Michnik, Józef Tischner, Jacek Żakowski (1995). *Między Panem a Plebanem / Between the Lord and the Priest*, Kraków: Wyd. Znak
- NKJP 50 = Zygmunt Kubiak (1989). *Przestrzeń dzieł wiecznych: eseje o tradycji kultury śródziemnomorskiej / The space of eternal works: essays on the tradition of Mediterranean culture*, Kraków: Wyd. Znak
- NKJP 51 = Jerzy Surdykowski (1989). *Duch Rzeczypospolitej / The spirit of the Republic*, Warsaw: PWN
- NKJP 52 = *Gazeta Wyborcza*, 30.08.1993
- NKJP 53 = *Gazeta Poznańska*, 2001
- NKJP 54 = *Gazeta Wyborcza*, 8.06.1993
- NKJP 55 = *Gazeta Ubezpieczeniowa*, 2007
- NKJP 56 = *Polityka* no. 2703, 2009
- NKJP 57 = *Gazeta Wyborcza*, 14.12.1992

- NKJP 58 = Gazeta Wyborcza, 7.02.1997
NKJP 59 = Polityka no. 2705, 2009
NKJP 60 = Słowo Polskie Gazeta Wrocławska, 17.09.2005
NKJP 61 = Gazeta Pomorska, 2010
NKJP 62 = Akant, 2008
NKJP 63 = www.forumowisko.pl 21.02.2006
NKJP 64 = Jacek Kwosek, Agnieszka Budzyńska-Daca (2009). *Erystyka czyli o sztuce prowadzenia sporów / Eristic, or the art of dispute resolution*, Warsaw: PWN
NKJP 65 = Polityka no. 2339, 2002
NKJP 66 = Polityka no. 2514, 2005
NKJP 67 = Gazeta Wyborcza, 25.04.1992
NKJP 68 = Polityka no. 2685, 2008
NKJP 69 = Gazeta Wrocławska, 31.12.2002
NKJP 70 = Krzysztof Burnetko, Witold Bereś (2008). *Marek Edelman. Życie. Po prostu / Marek Edelman. Life. Simply*, Warsaw: Świat Książki
NKJP 71 = Polityka no. 2651, 2008
NKJP 72 = Gazeta Wyborcza, 6.02.1999

Iwona Benenowska
ORCID iD: 0000-0003-1163-2778
Uniwersytet Kazimierza Wielkiego
w Bydgoszczy
Wydział Językoznawstwa
ul. Jagiellońska 11, p.5
85-067 Bydgoszcz
Poland
ibene@ukw.edu.pl

How much truth is there in *a true Pole*?

IWONA BENENOWSKA

*Received 17.08.2021,
received in revised form 14.11.2021,
accepted 10.01.2022.*

Abstract

The article is devoted to interpreting the image of *a true Pole* (re)constructed on the basis of the *National Corpus of the Polish Language* (NKJP) and Internet texts. The following issues were considered: stereotyping, evaluation, ideological thinking, profiling of terms, persuasive definitions, and approximation of meaning to answer the question contained in the title. The research has led to the conclusion that the concept of a true Pole can be considered on many research levels, and its use is associated with a specific ideological and/or axiological attitude.

Keywords

a true Pole, stereotype, evaluation, ideological thinking, concept profiling, persuasive definitions, meaning approximation

Ile jest prawdy w *prawdziwym Polaku*?

Abstrakt

Artykuł jest poświęcony interpretacji – (z)rekonstruowanego na podstawie *Narodowego Korpusu Języka Polskiego* (NKJP) i tekstów internetowych – obrazu *prawdziwego Polaka*. Aby odpowiedzieć na pytanie zawarte w tytule, wzięto pod uwagę następujące zagadnienia: stereotyp, wartościowanie, myślenie ideologiczne, profilowanie pojęć, definicje perswazyjne oraz aproksymację znaczenia. Badania doprowadziły do konkluzji, iż pojęcie *prawdziwego Polaka* można rozpatrywać na wielu płaszczyznach badawczych, a jego wykorzystanie wiąże się z określona postawą ideologiczną i/lub aksjologiczną.

Słowa kluczowe

prawdziwy Polak, stereotyp, wartościowanie, myślenie ideologiczne, profilowanie pojęć, definicje perswazyjne, aproksymacja znaczenia

1. Introduction

The aim of this text is an attempt to interpret the (re)constructed image of a *true Pole*¹ and to verify the following research hypothesis: the notion of a *true Pole* undergoes specific changes (within the framework of the so-called redefinition of the stereotype), is used in discussions based on ideology and axiology, and simultaneously belongs to the research planes of several disciplines.

The image of a contemporary *true Pole*, which has been reconstructed based on almost a hundred texts from the *National Corpus of the Polish Language* and selected statements by Internet users, shows a polarised image.

Let us briefly recall that a person defined by this term is: a Catholic, a nationalist, a patriot, holding Polish citizenship, recognising Polish as the language of the national community,

¹ See the article entitled *Próba rekonstrukcji językowego obrazu prawdziwego Polaka (Attempt to reconstruct the linguistic picture of a true Pole)*.

hospitable, chivalrous, family-oriented, but also an opponent of evolutionism, socialism/communism, Nazism, liberalism; a nationalist convinced of the existence of enemies of Poland, prejudiced against representatives of certain nations, perceiving cultural and political threats coming from abroad; a person with a taste for coarse jokes, usually behaving too loudly or in an inelegant and uncultured manner, and moreover, stubborn (aggressive) in defence of his or her opinions, with complexes (also in the scope of lack of knowledge of foreign languages), prone to drunkenness, criticism, complaining, backbiting, indolence, intolerance, envy, egoism, jealousy, hypocrisy, offence, etc.

Taking the above into account, it is worth considering how such an image functions and with what it is potentially associated.

2. Interpretative categories and comments

Based on the review of the features included in the notions of *a true Pole* and *nation*, one can notice a convergence of explanations, especially on the cultural (cultural heritage, language, traditions, customs) or ideological (religion, patriotism, national symbols) plane.

Nonetheless, what is the reason for the **polarisation** of valued characteristics? What reflections arise in connection with the construction of the image of *a true Pole*?

2.1. Stereotype

Suppose we assume that the stereotypes are (Jerzy Bartmiński and Jolanta Panasiuk's definition) a set of particular features abstracted from the texts. In that case, they confirm the fixation of these in the consciousness of language users in a specific meaning and carry specific connotations:

they confirm consolidating certain features in the colloquial characteristics of the objects in question and including the features in the linguistic image of the objects, in the connotation of the words

that name them. [...] Sets of these features, internally organised in a non-accidental way, create linguistic and cultural images of objects, which since Walter Lippmann have been called stereotypes (Bartmiński, Panasiuk 1993: 363), a true Pole is an example of a classic stereotype. The above-mentioned researchers presented a set of characteristics resulting from ASA'90 research (ibidem: 380)².

In the analysed material concerning *a true Pole*, the variants of the stereotype: pattern (*true*) and image (*typical*) are mixed, resulting from the fact that the language users in their texts assign the features without the key as mentioned above. These are spontaneous statements, not surveys.

Extralinguistic knowledge, which is drawn from one's own experience, and dictionary knowledge recorded in dictionaries interact with each other, resulting in a specific linguistic and cultural image of the named objects. Hence:

The object (subject) P (e.g., *a Pole...*) of x characteristics (a member of the nation, using Polish as the native language...), called with lexemes L_1 (*a patriot...*) or L_2 (*a xenophobe...*) having different linguistic connotations 'y' ('a man loving his homeland') and 'z' ('a man reluctant or hostile towards foreigners') receives a variant stereotype (S_1, S_2) of partially differentiated content. The differentiation is connected to subjective evaluation.

2.2. Evaluation

Ambivalent connotationally and lexically evaluative judgements were noted in the group of indicated features. Let us use only a few examples, which will make the mechanism of reversing the sign of valuation depending on the judgments of the subject of valuation explicit:

A Catholic with his worldview, moral and social attitude connected with the adherence to Christian principles, which are

² I write on this subject in detail in the article entitled *Próba rekonstrukcji językowego obrazu prawdziwego Polaka (Attempt to reconstruct the linguistic picture of a true Pole)*.

associated with the love of God and neighbour, mercy, striving for peace, and so forth, connotes positive valuation. He/she is a bearer of positive values, especially social ones (but also sacral – striving for sanctity, ethical – love of the truth), which is closed in the judgment formula:

X (a Catholic) is a D-bearer of W social, sacral, ethical

Such evaluation is characteristic for our cultural circle, i.e., the subject (of valuing) is usually a member of the community and recognises similar values.

In the collected material, there were examples in which a subject from the same cultural circle (also a Pole) confronts positive connotations with his/her own (negative) ones, i.e., the commonly accepted image of a Catholic has been eroded, specific actions, which are negatively evaluated, are indicated e.g.

[...] in my opinion, a "**true Pole**"³ is the one who **listens to a priest** every Sunday,⁴ and instead of using his/her reason and simply thinking independently, he/she lets himself/herself **be led like a silly little sheep** by its shepherd... [net. 1].

That is: human's actions (*he/she listens to the priest* = obeys him, *lets himself/herself be led*) are cognitively wrong since they deprive the object of independent thinking in favour of blind trust in the clergyman. The formula of an evaluative judgement has the form:

X (the fact that he/she listens to the priest, lets himself/herself be led) is Z W cognitive

³ Original spelling has been retained in quoted passages.

⁴ There is the author's emphasis throughout the text.

In the example:

[...] **a true Pole** (or elsewhere: a true Irishman, Spaniard, etc.) can only be one who **believes in God** (net. 2);

it is about potentially generating conflicts, starting wars in the name of professed faith, which is ethically wrong. It can be clearly seen in the rest of the quotation:

In the fight for Poland (whether in defence against an invader or in an arch-righteous assault in the name of Polish super power status), it translates into the slogan "Beat whoever believes in God/ Beat who believes in god". Beat who? The one who does not believe in God – an apostate, a heretic, a blasphemer, even more so an unbeliever, and certainly a godless person, who by his/her very existence offends God (net. 2).

The subject's value judgement⁵ closes in the formula:

X (the fact that he/she kills in the name of God, sees an enemy in the follower of another faith) is Z W ethical.

Noteworthy are the cases in which the author uses irony. It is deciphered on the basis of the context, or marked in the text by putting a given word in inverted commas, e.g.:

A true Catholic Pole would never accept that the bard Adam Mickiewicz, Jesus Christ and the Queen of the Polish Crown were really Jews. – Jews are few in Poland, but the attitude to Jews is as far from indifference as before the war [...] (NKJP 1)

Irony permanently reverses the mark of evaluation, and in the above example, the Catholic is valued negatively for his anti-Semitism. Such a Catholic is a non-Catholic (a human who

⁵ At the same time, if the subject were someone sharing the point of view of a true Pole, the sign of valuation would be reversed: X (the fact that he/she kills in the name of God, sees an enemy in the follower of another faith) is D W ethical.

loves his/her neighbour), so the following formula of value judgment can be proposed:

X (a Catholic who hates Jews) is Z-bearer W ethical

Furthermore, it cannot be excluded that the sender of the value judgement (the subject) represents another religion and/or belongs to another cultural circle. The negative evaluation is then also possible. Someone who is a Catholic becomes the enemy – the bearer of bad qualities.

Names of such understood values then become a label, sometimes even a stigma, which serves to identify in the relation "one's own – stranger", and in the other sequence unethical exclusion.

Similar relationships are noticeable in other names of carriers of qualities, e.g., a patriot, a nationalist.

Among the names of the bearers of negative features there are lexemes not only connoting negative evaluation, but also containing valuing elements in their semantic structure (the so-called semantic valuing), e.g., a xenophobe ('a human reluctantly or hostile disposed towards foreigners and foreigners' USJP), an anti-Semite ('an opponent, enemy of Jews' USJP), drunkard ('one who compulsively drinks alcohol, who gets drunk; an alcoholic' USJP) etc.

To sum up this part, the negative evaluation of a true Pole is a much more elaborate area and entangled in contexts than the positive evaluation (given explicitly). It may involve the so-called ideological thinking.

2.3. Ideological thinking

Evaluation is a process characterised by subjectivity, subordinated to certain tastes or interests. Besides, as Mateusz Witkowski writes:

'Binary divisions such as "one's own – stranger", "we – they", "true – false" are, of course, the salt of any **propaganda** activity. Leaving

aside the aforementioned lists: we are constantly bombarded with opinions in which the representatives of various political fronts reach for the category of "truthfulness". Who, then, is this true Pole, the opposite of a spy and a dyed fox? This, of course, depends on the speaker's perspective. (net. 3)

Jadwiga Puzynina writes about the relation between values and ideology, pointing out that the notion of value is fundamental "for all ideologies, always based on certain values" (2008b: 17). It is critical to grasp the relationship of their interpenetration from the subject's perspective. An ideology is subject to evolution, and an equally important feature is a social group's adherence to it. At the same time, "the primary, defining subject is the social group, and the secondary subject is the individual who belongs to it, joins it, or even creates it. As long as some views are only proper to it, we will not call it an *ideology*" (Puzynina 2008b: 17).

The evolving image of *a true Pole*, identification with the features ascribed to him/her both by an individual and a community/social group, situates this concept in the field of ideologies, to be more precise – symbolic ideologies understood as mental schemes strongly influenced by emotions, and cognitively deficient, which are signed by individuals identifying themselves with them (Skarżyńska 2005, after Puzynina 2008b: 19).

The changes in the concept of *a true Pole* are illustrated by the process of the so-called ideological thinking, about which Mirosław Karwat writes as follows:

Ideological thinking is "thinking according to values". Emotions ascribed to specific values dominate here over reflection and critical analysis of phenomena. The desire for objectivity, for distance, does not arise here; it is replaced by a sense of the obviousness of views (perceptions and judgements) born of prejudice. In the subject's consciousness and in his/her way of communicating with others, the difference between the language of description, the language of values (judgements, patterns, ideals), and the language of norms (orders, prohibitions) is blurred. Social phenomena are perceived and presented not as such, but are immediately marked with

a mark of evaluation – positive or negative. [...] The consequence is a programmatic bias of interests and perceptions, selectivity of the image of phenomena [...], biased perceptions and messages addressed to others [...] (2008: 42).

We know from observation (and autopsy) how difficult it is to discuss with a person representing ideological thinking as he/she always knows better, adheres to the principle *my truth is "mine"* (= better), and confrontation with another point of view generates bad emotions and does not promise a factual discussion, but conflict, as in the examined material, e.g.:

In the previous years, depending on the needs and what they preached, the **most enlightened men of the Polish Church were called the Jews for beating**: Father Professor Józef Tischner, Bishop Józef Życiński, Cardinal Franciszek Macharski and Pope John Paul II. However, a **true Pole, Catholic, has never accepted the fact** that the bard Adam Mickiewicz, Jesus Christ and the Queen of the Polish Crown **were really Jews**. (NKJP 2)

You Israeli stooge! Mazowiecki can rule in the synagogue, not in Poland! (NKJP 3)

I'm not going to respect someone who hurts me – **because I've already once been ripped off by one who called himself a "real Pole"** and added "not to mess with the skins". (NKJP 4)

2.4. Profiling of concepts

The configuration of features in the image of *a true Pole* is subject to change, some withdraw, and others take their place. Their arrangement is motivated by individual needs, experiences, and values. In keeping a set of the same (or similar) traits, their internal structure is rebuilt, and **the rank** of individual aspects to which the traits refer changes. Moreover, there are shifts on the scale of *good – bad / plus – minus*. It affects the modification of meaning, the organisation of the components of meaning, revealing at the same time certain truths about the

subject and how he/she organises judgements about reality and its conceptualisation. The base idea is modified, in other words, it is subjected to a process called **concept profiling**, and this means that it has occurred:

a subjective (i.e., having its own subject) linguistic-conceptual operation consisting in the specific shaping of the image of an object through its inclusion in certain aspects (subcategories, facets), [...] within the framework of a certain type of knowledge and according to the requirements of a specific point of view (Bartmiński, Niebrzegowska-Bartmińska, Nycz 2004: 212).⁶

Profiling is the domain of discourse. The discourse assumes the presentation of someone's statements intentionally aimed at achieving specific communicative goals. For this reason, profiling is a tool frequently applied in ideological, political, and social disputes. At the same time, I adopt the way of understanding and the model of description present in Jerzy Bartmiński's works (and other relevant publications), i.e., conducted on two concretisation levels: a given (concrete) utterance and a social convention (with the possibility of repetition in various utterances), and anthropological, focused around the speaker, representing the subjective type of rationality of the sender of information in relation to his/her knowledge about the world and the system of values. The interpretation manner would also assume naming the elements constituting the image of *a true Pole* as facets, discovered through the analysis of the material (cf. Bartmiński, Niebrzegowska 1998: 214). The model was successfully used by Bartmiński while describing the changes in the stereotype of the German in Polish (Bartmiński 1998: 225-235). Apart from the changes within the bundle of features, one can see profiling at the level of the sender – the speaking subject. From the point of view of a particular worldview, negative features and valuations predominate in the texts of, e.g., *Gazeta*

⁶ Concept profiling has no single/unique definition. There are also at least two ways to apply the concept. I omit the one derived from American cognitivism.

Wyborcza (statements such as: *And finally, a unique candidate: he does strange business, goes bankrupt and comes out on top, drinks, swears, defends the Church, does not like Jews – in a word a 'real' Pole* (NKJP 5)), Polityka (e.g., *Against the Saxons and the Swedes, also against Russia. In short, against the whole world. A true Pole* (NKJP 6)), Newsweek (e.g., the very title of an article by Włodzimierz Mędrzecki: *Luters, Dandies, Traitors, Jews. Who can and who should not call themselves a “true Pole”* (net. 4), the OKO.press portal (e.g., “According to Kaczyński, Morawiecki and co., only those who support their politics and conservative, Catholic ideology are Poles. Who does not support them is not a Pole, and in a more robust version is a German. [...]” (net. 5)) – sources considered to be left-wing, liberal.⁷

2.5. Persuasive definitions

When emotional states are replaced or mistakenly interpreted as cognitive processes, we may be dealing with a persuasive situation. It is especially true for terms that are "burdened" with conceptual ambiguity and are emotionally neutral. It is then that an attempt may be made to specify or define them: *What is X?*, and create a persuasive definition. *A true Pole* is one of them. As Jakub Pruś writes:

[...] the evaluative term 'true' etc., probably due to its normative character, plays a different role than the traditional differentia specifica; its task is to influence the original meaning of the term so that it takes on the desired scope for further persuasion. The same is true of examples such as a reliable scientist or a normal man, or an authentic Christian – which Christian would want to be inauthentic, which normal man wants to be abnormal, and which scientist wants to be unreliable? (Pruś, 2019: 62)

The above judgment is also shared by other users of Polish. In the collected material, observations regarding the motivation for

⁷ It is based on observations of the collected material, but the issue requires separate research on more extensive source material.

the use of the expression in question were noted. For example, Kazimierz Piotrowski in a conversation with Jarosław Suchan openly signalled:

[...] the expression a True Pole is a **rhetorical expression**, not a descriptive one, because there are no true Poles, just as there are no untrue Poles. The point is to be fully aware of how we use this expression and for what purpose we do it (NKJP 7)].

An internet user (nicknamed⁸ Antoni222 Sułek in the article *Instrukcja dla niezdecydowanych wyborców (Instructions for undecided voters)*) notes:

Values can be implemented variously and no one is a "true" Catholic, a "true" Pole or a "true" Populist. **This is the sand that politicians pour into your eyes** so that you cannot see their true face or the true misery of their programmes (NKJP 8).

For those reasons, the composed bundle of features and the ways of clarifying/defining⁹ them in the image of a true Pole resemble the construction of a persuasive definition. It is defined in the *Encyklopedia PWN* as:

a verbal statement in the form of a definition, in which the choice of expressions is determined by the intention to evoke (or change) a certain emotional attitude (belief, decision) (net. 6).

The intention of its use is emphasised by Charles L. Stevenson, writing that a persuasive definition gives a known word a new conceptual content but does not change its emotive meaning much, and it is applied (usually consciously) to change the

⁸ Nickname is used in Polish language programs to denote a pseudonym, login, identifier, etc.

⁹ It is easy to confuse the construction of a persuasive definition with an attempt at clarification. A differentiating factor, i.e., the sender's intention, comes into play. If the intention of the language user is to change attitudes and judgements in the recipient, rather than a desire to gather facts, then we have an example of a persuasive definition.

listeners' preferences (2015: 21).¹⁰ It should be kept in mind that the clearer the emotive meaning becomes, the more turbid the conceptual content of the word may be. As a result, emotional states are misinterpreted as cognitive processes (*ibidem*). Such a construction aims to redirect the preferences of language users, i.e., change the (degree of) acceptance of the described phenomenon. Thanks to the afore-mentioned opacity of the content, the change is felt as 'natural', and its results are easily fixed. In this way, the recipient is dissuaded from uncritical approval of a specific set of qualities while being induced to praise others, so their conceptual content undergoes a continuous **re-definition process** (2015: 21). Stevenson hints that persuasive definitions are often identifiable by the appearance of phrases such as 'in fact' or 'true'/'really' in a metaphorical sense, in the examples of charity, love, courage, where "charity" is not giving gold/money but understanding, true love is a communion of minds, and courage in everyday reality "defies public opinion, where the expressions of 'true' gain the persuasive power of 'being accepted'" (2015: 21).

Let us look at selected examples taken from the linguistic material studied. We will take one feature of a true Pole – patriotism. Stanisław Berenda-Czajkowski in his 2001 novel *Dni grozy i łez (Days of horror and tears)* writes:

[...] every true Pole must be **a soldier. Faithfulness** to the tormented homeland, "at any time and any day" [...]. (NKJP 9)

The left-wing¹¹ *Dziennik Trybuna* (Daily Tribune) of 28 June 2016, on the other hand, reads:

¹⁰ Stevenson is considered a precursor. In Polish research, Tadeusz Pawłowski developed the concept by presenting three types of persuasive definitions: 1. to change the scope of the definiendum, 2. to change the emotions, 3. to replace a previously used concept by another concept with a different emotional charge (1978: 228-238). Grażyna Solecka (2010) and Jakub Prus (2019) have also addressed this topic.

¹¹ See the term at <https://wiadomosci.dziennik.pl/media/artykuly/449630,dziennik-trybuna-zniknie-zespol-odchodzi-z-redakcji-przez-zwolnienia-i-ciecicia-pensji.html> [Accessed: 24.09.2020].

The most common variety of **patriotism assumes unconditional love for the homeland**, which is **understood sentimentally** as a familiar landscape, original cuisine, folk customs, etc., and ultimately the place where our ancestors were born and lived. This sentiment is guarded by the collective of true Poles, i.e., the ethnic element. [...] However, more interesting is **the pro-state patriotism**, which has a very short history in Poland. Representatives of this trend do not see Polishness in the graves of their ancestors, in the landscapes on the Vistula or in the aroma of bigos, but, for example... in **paying taxes**, which are getting increasingly expensive every year, or in **paying the bus fare**. (net. 7)

There is a clear redirection of the viewer's attention to a bundle of qualities other than the traditional (established) ones: the relevant contemporary qualities are paying taxes, validating tickets, etc., activities aimed at not stealing from the state, while it seems that the state is understood as a community – all Poles and their welfare.

Tadeusz Kwiatkowski also points to the possibility of changing the direction of evaluation or the content of entire value judgements, which is the domain of persuasive quasi-definitions (Kwiatkowski 2002: 402), i.e., procedures which enable re-evaluation. The re-evaluation, in turn, may consist in blurring and changing the evaluative meaning of the term being defined (cf. Pietrzak 2015: 62 et al.). A good example is the modifications related to the word *friend*. The word 'concubine' (to use Stevenson's example, 2015: 26) has become to some extent vulgar and unpleasant to our ears. The word 'friend', displacing the previous one, provided a convenient ambiguity as it partly retained its former positive meaning and thus became helpful in taking off the social odium from concubines.

Among the characteristics of *a true Pole* discussed in this text, a good exemplification material is to understand the word *a liberal*. In dictionaries, we find that it is a supporter of liberalism, an advocate of tolerance towards disapproved views, attitudes and other people's deeds or a political direction advocating the guarantee of individual freedom, national and religious minorities, etc. (cf. USJP). Such a meaning typically connotes

a positive valuation of the bearer of the feature named with this lexeme. On the other hand, Piotr Cieśliński in his article *Między nowym totalitaryzmem a nihilizmem aksjologicznym (Between new totalitarianism and axiological nihilism)* in *Gazeta Wyborcza* of 7.02.1997 indicates the possibility of extracting a different meaning by a specific type of language use and the reversal of the sign of evaluation:

Today 'we' are the true Poles and 'you' are liberals, enemies of Poland or even Europeans. (NKJP 10)

The liberal thus (re)defined is already evaluated negatively. It is accompanied by an extension of the negative valuation to the words constituting the immediate context: *Europeans*. One can clearly notice the manipulative potential of quasi-persuasive definitions, which Teresa Hołówka writes about:

They manipulate the meaning of a term, usually a not very clear, ambiguous term, lacking clear, paradigmatic uses, while creating the impression that either one is merely describing the accepted usage or revealing the 'heart of the matter' hidden in it. (2012: 61).

Finally, it is worth mentioning that Ewa Masłowska describes the phenomenon of changing the meanings of certain words under the influence of the value system of a given social group using the concept of *secondary meanings* (1991: 181-185). Thus, we come to the conceptual category of *truth*.

2.6. Approximation of meaning

Many texts have been devoted to the truth as a conceptual category (mainly philosophical). Among others, Jadwiga Puzynina (2008a) wrote about the importance of truth in linguistic communication, as well as about the multiple meanings of the word *truth*; Aleksander Kiklewicz (2017) mentioned the interpretative aspect of meaning and the understanding of content as a condition for the realization of the category of truth; Dwight Bolin-

ger (1973) stressed that truth constitutes the norm of linguistic communication, while Robert Pilat considers it the basis of all action (2009), and – as we well remember – Herbert P. Grice – as a condition for implementing the communicative principle of cooperation in terms of the category of quality:

"Try to make your participation such that it is truthful" and two more specific maxims:

1. Do not say what you think is false.
2. Do not say what you do not have due justification for. (1977: 89)

Meanwhile, Kiklewicz states:

However, the linguistic system works algorithmically, allowing the production (or reproduction) of constructions (messages) **useful for human** (especially **interactional**) activity regardless of whether they are true or false, appropriate or inappropriate, acceptable or unacceptable. [...] Since we can understand these expressions, they fulfil their function, while the fact that they are true or false does not impinge on the language system or the linguistic competence of its users (2017: 7).

A true Pole and the differences in understanding/defining it resemble the issues of correct/incorrect referentiality in the case of *nomina appellativa*, where referentiality can be quite complicated. It is, among other things, due to names with an undefined and subjectively conditioned conceptual content, such as *a banana* that has a fixed and defined conceptual scope, while *democracy* has a fuzzy scope as it depends to a large extent on the subjects' attitudes, political views, and shared ideologies (cf. *ibidem* 26). Our cognitive system is thus a kind of **medium for veridical interpretation**, because:

Truth does not mean the direct conformity of a sentence to **reality**, but conformity to the **image of reality** cultivated by communities, groups, or individuals. The recognition of anything as true or false is essentially based on our beliefs, partially verified in practice. Since beliefs belong to categories fixed in consciousness, they are

often overlooked in reasoning processes. We consider something to be true or false, not being fully aware of the fact that it is true or false for us, according to us, from our point of view (ibidem 34).

Thus, from this perspective, the image of *a true Pole* appears as an approximation of the truth, according to Kiklewicz's definition:

Approximation (or semantic diffusion) means underdetermination, underspecification of the meaning of linguistic units (words, sentences, texts), the blurred nature of the boundaries of lexical concepts, i.e., the lack of conviction of language users as to what content belongs to the scope of the unit's meaning and what not (2020: 10).¹²

Our extra-linguistic knowledge gives grounds to avoid large quantifiers as we realise that the sentence: *All true Poles/every true Pole is like that ...* (+ a bunch of features) is false, but true or at least highly probable is: *There are some true Poles who*

3. Conclusion

The source of shaping the characteristics of a young true Pole may be patterns drawn from the surrounding reality, about which Jolanta Saacewicz writes as follows:

It is enough that he/she goes to a Polish school. It is enough that he/she sits in front of TVP, that he/she is surrounded by Catholic community and led by a real Polish priest, that he/she sees those monuments leading to heaven, that he/she soaks in the legend of 'Bury' and the Świętokrzyska Brigade, that he/she listens to Radio Maryja, that he/she hums disco polo and goes to the cinema to see 'Smolensk'! It is enough for him/her to be surrounded by Poland (net. 8).

¹² In his monograph, the author gives examples of various types of approximation and their motivation, e.g., approximation of derivatives, names, axiomatisation, semantic phantoms, and many others (Kiklewicz 2020: 10-90).

In the summary of the previously quoted text, Jakub Żulczyk rightly notes that the axiologically polarised compositional elements of the image of a true Pole are:

[...] simply a clumsy attempt to catch the elusive, i.e., 'Polishness', that which is Polish, which can describe our merry nation. [...] a true Pole is someone who works hard while earning far too little. He/she is someone confused, distrustful and suspicious, because the state in which he/she lives guarantees him/her absolutely nothing, except another heist on his/her hard-earned change. I would also say that this is someone brave. Accustomed to poverty. Reasonable. In fact, tolerant, kind, and witty. Honest and articulate. In fact, whatever happens, I have in my heart the conviction that if there is a formula for a true Pole, it includes a kind of, broadly defined, being right, despite extremely unfriendly circumstances. And the conviction that it will be all right after all. All in all, I have the impression that this conviction is in us. After all, we are all True Poles. We all have our ghastly abstract language and Polish identity cards. And above all, we were born here. No one made us from a rubber-like product in a Chinese factory near Rzeszów (net. 9).

In her monograph entitled *Wśród stereotypów i tekstów kultury* (*Among stereotypes and cultural texts*), Małgorzata Karwatowska draws certain conclusions that are also relevant to the notion referred to in this text, namely, writing about **the stereotype**. She notes that it assumes the character of a label, provides the basis for deriving specific evaluations, facilitates the diagnosis of phenomena, contributes to the creation of social bonds, and constitutes a set of judgements about a part of reality (2020: 17-18).

Simultaneously, in the last twenty years (the turn of the 20th and beginning of the 21st centuries), there has been a clear tendency to redefine the notion along with the changing extra-linguistic reality, the dynamics of political and social transformations. Jolanta Panasiuk made similar observations about each stereotype:

[...] along with changes in language, changes in extra-linguistic reality, stereotypes may change, while the quality and direction of the changes need not be parallel (1998: 97).

When writing about moral criteria of contemporary discourse, Mieczysław Michalik takes up the issue of generalising, “juggling” facts depending on non-communicative aims. Therefore, the notion of *a true Pole* will be treated instrumentally as an element of **moral demagogy**:

In discussions about social reality, individual phenomena and facts are extremely often treated as general tendencies, raised to the rank of universal phenomena – a part taken for the whole. Almost every thesis can be “proved” that way, and every thesis can be supported by at least one fact, which is considered an exception confirming the rule. Such reasoning proceeds here according to the following scheme: “X is a thief (possibly also Y and Z), so all people (of a given community, group) are thieves” (1980: 164-165).

or moral **blackmail** or **discredit**:

It always serves to increase the emotional pressure on the recipients of such statements. It is also connected with moral blackmail – if someone does not support my demands, calls, warnings ..., they become morally suspect (Karwat 2006: 40).

The expression a “*true Pole*” may also be used willingly by someone who aims at discrediting the Other by attributing certain features to him/her or refusing to acknowledge other features,¹³ a populist (when he/she mythologises the common wisdom of the “simple human” to gain supporters) or a **demagogue**, when:

appears from the position of “spokesperson and defender of the people”. If he/she emerges not based on 'one of us' [...], it takes place on the principle of a “true Pole” (respectively: a true Catholic, a com-

¹³ Mirosław Karwat writes more about various mechanisms of discrediting in the chapter *Plaszczyzny i kryteria dyskredytacji (Discrediting planes and criteria)* (2000: 135-154).

munist, an anti-communist), who warns against those untrue, tracks them down, exposes, and stigmatises them (Karwat 2006: 36-37).

The research conducted so far shows that the notion of a true Pole, which is standardly included in the stereotype, entails various interpretative possibilities. It shows features of profiling notions and clear definitions; it is an effect of ideological thinking and could be a tool of unethical manipulation, discrediting, while the category of truth has only the character of approximation. Therefore, it can be studied and described in various categories belonging to linguistics, sociology, and philosophy. And with the knowledge of other contexts – also another reinterpretation. If it is referred to in one, it gains a greater degree of approximation.

Finally, a reflection arises that when observing the use of a notion, we will not get much truth about it but a lot about its users, who often share the judgment: We consider as reasonable only those who are of our opinion.

References

- Bartmiński, Jerzy (1998). "Zmiany stereotypu Niemca w Polsce. Profile i ich historyczno-kulturowe uwarunkowania" In: Jerzy Bartmiński, Rafał Tokarski (eds.). *Profilowanie w języku i w tekście*. Lublin: Wydawnictwo UMCS, 225-235.
- Bartmiński, Jerzy, Stanisława Niebrzegowska (1998). "Profile a podmiotowa interpretacja świata" In: Jerzy Bartmiński, Rafał Tokarski (eds.). *Profilowanie w języku i w tekście*. Lublin: Wydawnictwo UMCS, 211-224.
- Bartmiński, Jerzy, Jolanta Panasiuk (1993). "Stereotypy językowe". In: Jerzy Bartmiński (ed.). *Encyklopedia kultury polskiej XX wieku*. Vol. 2. *Współczesny język polski*. Wrocław: Wiedza o Kulturze, 363-388.
- Bartmiński, Jerzy, Stanisława Niebrzegowska-Bartmińska, Ryszard Nycz (2004). *Punkt widzenia w tekście i w dyskursie*. Lublin: Wydawnictwo UMCS.
- Bolinger, Dwight (1973). "Truth in a Language Question". *Language* 49/3: 539-550.

- Dubisz, Stanisław (ed.) (2005). *Uniwersalny słownik języka polskiego (USJP)*. Warszawa: PWN.
- Grice, Herbert Paul (1977). "Logika a konwersacja". Trans. Jadwiga Wajszczuk. *Przegląd Humanistyczny* 6/141: 85-99.
- Hołówka, Teresa (2012). *Kultura logiczna w przykładach*. Warszawa: PWN.
- Karwat, Mirosław (2008). "Cechy myślenia ideologicznego". In: Irena Kamińska-Szmaj, Tomasz Piekot, Marcin Poprawa (eds.). *Ideologie w słowach i obrazach*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 42-49.
- Karwat, Mirosław (2006). *O demagogii*. Warszawa: Wyd. Sejmowe.
- Karwatowska, Małgorzata, Leszek Tymiakin (2020), *Wśród stereotypów i tekstów kultury. Studia lingwistyczne*. Lublin: Wydawnictwo UMCS.
- Kiklewicz, Aleksander (2020). *Fenomeny komunikacji (normy i dewiacje w zachowaniach językowych)*. Olsztyn: Centrum Badań Europy Wschodniej.
- Kiklewicz, Aleksander (2017). *Znaczenie a prawda. Fantomy semantyczne*. Olsztyn: Centrum Badań Europy Wschodniej.
- Kwiatkowski, Tadeusz (2002). *Wykłady i szkice z logiki ogólnej*. Lublin: Wydawnictwo UMCS.
- Masłowska, Ewa (1991). "Kształtowanie się wtórnych znaczeń wyrazów pod wpływem obowiązującego w danym społeczeństwie systemu wartości" In: Jadwiga Puzynina, Jerzy Bartmiński (eds.). *Język a Kultura*. Vol. 2. *Zagadnienia leksykalne i aksjologiczne*, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 181-185.
- Michalik, Mieczysław. (1980). *Szkice o kryteriach moralnych*. Warszawa: Wiedza Powszechna.
- Panasiuk, Jolanta. (1998). "O zmienności stereotypów". In: Janusz Anusiewicz, Jerzy Bartmiński (eds.). *Język a Kultura*. Vol. 12. *Stereotyp jako przedmiot lingwistyki*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 84-98.
- Pawłowski, Tadeusz. (1978). *Tworzenie pojęć i definiowanie w naukach humanistycznych*. Warszawa: PWN.
- Pietrzak, Marcin. (2015). "Pojęcia definicji i guasi-definicji perswazyjnej". *Ruch Filozoficzny* LXXI/3: 53-68.
- Piłat, Robert. (2009). "Prawda jako wewnętrzna norma sądu. Husserl wobec analitycznej teorii prawdy". *Przegląd Filozoficzny. Nowa seria* 18/4: 487-507.
- Prus Jakub. (2019). "Argumentacja semantyczna – podstawowe pojęcia i problemy definicyjne". *Res Rhetorica* 6/4: 56-76.

- Puzynina, Jadwiga. (2008a). "Co się dzieje z prawdą dziś?". In: Anna Dąbrowska (ed.). *Język a Kultura*. Vol. 20. *Tom jubileuszowy*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 35-49.
- Puzynina, Jadwiga. (2008b). "Ideologia w języku polskim". In: Irena Kamińska-Szmaj, Tomasz Piekot, Marcin Poprawa (eds.). *Ideologie w słowach i obrazach*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 13-23.
- Skarżyńska, Krystyna. (2005). *Człowiek a polityka. Zarys psychologii politycznej*. Warszawa: Wydawnictwo Naukowe Scholar.
- Solecka, Grażyna. (2010). "Dyskursywny charakter definicji". In: Wojciech Suchoń, Irena Trzcieniecka-Schneider, Dominik Kowalski (eds.). *Argumentacja i racjonalna zmiana przekonań*. Kraków: Wydawnictwo UJ, 45-51.
- Stevenson, Charles Leslie. (2015). "Definicje perswazyjne". Trans. Marcin Pietrzak. *Studia z Historii Filozofii* 2/6: 21-43.

Netography

- net. 1 = <http://joanna-kaffe-latte.blogspot.com/2015/10/prawdziwi-polacy.html> blogspot Joanna – Świat według mnie / Joanna – *The world according to me*, 17.10.2015 [Accessed: 16.09.2020]
- net. 2 = <https://trybuna.info/opinie/bog-honor-dziczyzna/> [no author], *Bóg, honor, dziczyzna / God, honour, game*, Dziennik Trybuna 21.06.2017 [Accessed: 16.09.2020]
- net. 3 = <https://ksiazki.wp.pl/szczucie-zydem-i-moralnosc-kalego-6145577583786113g/5> Mateusz Witkowski, *Szczucie Żydem i moralność Kalego / Joking with a Jew and Kali's morality*, portal WP.pl – reference, among others, to the issue of *Zła mowa / Bad speech* M. Głowińskiego, 16.11.2016 [Accessed: 16.09.2020]
- net. 4 = <https://www.newsweek.pl/wiedza/historia/polska-tozsamosc-kto-moze-nazywac-sie-prawdziwym-polakiem/ee0w4ls> Włodzimirz Mędrzecki, *Lutry, fircyki, sprzedawczyki, Żydy. Kto może, a kto nie powinien mienić się „prawdziwym Polakiem” / Luters, Dandies, Traitors, Jews. Who can and who should not call themselves a "true Pole"*, Newsweek, 19.10.2016 [Accessed: 16.09.2020]
- net. 5 = <https://oko.press/jeden-narod-jeden-kosciol-jeden-kaczynski/> Piotr Pacewicz, *Jeden Naród, jeden Kościół, jeden Kaczyński / One Nation, one Church, one Kaczyński*, Oko.press 21.07.2020 [Accessed: 16.09.2020]

- net. 6 = <https://encyklopedia.pwn.pl/haslo/definicja-perswazyjna;3956098.html>] Encyklopedia PWN
- net. 7 = <https://trybuna.info/opinie/nie-prawica-nie-lewica-tylko-indywidualizm/> [no author] *Nie prawica, nie lewica, tylko indywidualizm? / Not the right, not the left, but individualism?* Dziennik Trybuna 28.06.2016 [Accessed: 16.09.2020]
- net. 8 = <https://wyborcza.pl/7,162657,25960327,kazdy-prawdziwy-polak-to-skarb-narodowy-i-pomnazanie-narodu.html> Jolanta Saacewicz, *Każdy prawdziwy Polak to skarb narodowy / Every true Pole is a national treasure*, Gazeta Wyborcza 21.05.2020 [Accessed: 16.09.2020]
- net. 9 = <https://facet.onet.pl/warto-wiedziec/po-czym-rozpoznać-prawdziwego-polaka/t7lvqn1> Jakub Żulczyk, *Po czym rozpoznać prawdziwego Polaka? / How to recognise a true Pole?* – Onet 25.04.2013 [Accessed: 16.09.2020]

NKJP – *Narodowy korpus języka polskiego (National corpus of the Polish language)* http://nkjp.uni.lodz.pl/ParagraphMetadata?pid=32dc5f02515f8f99577b6dbf90c7be3c&match_start=423&match_end=438&wynik=36#the_match [Accessed: 10.07.2020]

NKJP 1 = Jerzy Sławomir Mac, Wprost No. 46, 1997

NKJP 2 = Jerzy Sławomir Mac, Wprost No. 46, 1997

NKJP 3 = Gazeta Wyborcza, 26.08.1995

NKJP 4 = www.forumowisko.pl 21.02.2006

NKJP 5 = Gazeta Wyborcza 2. 11.1995

NKJP 6 = Polityka No. 2484, 2004

NKJP 7 = Rozmowa: Jarosław Suchan, Kazimierz Piotrowski, Onet.pl 24.07.2003

NKJP 8 = Gazeta Wyborcza, 11.09.1993

NKJP 9 = Stanisław Berenda-Czajkowski, *Dni grozy i łez / Days of horror and tears*, Wyd. Czytelnik, 2001

NKJP 10 = Gazeta Wyborcza, 7.02.1997

Iwona Benenowska
ORCID iD: 0000-0003-1163-2778
Uniwersytet Kazimierza Wielkiego
w Bydgoszczy
Wydział Językoznawstwa
ul. Jagiellońska 11, p.5,
85-067 Bydgoszcz
Poland
ibene@ukw.edu.pl

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

<https://doi.org/10.26881/bp.2022.1.06>

**Forms of address and (im)politeness:
A corpus-assisted study of Polish professional
and non-professional subtitles**

ANNA BĄCZKOWSKA

*Received 11.10.2021,
received in revised form 22.02.2022,
accepted 20.12.2022.*

Abstract

The aim of the study is to analyse how the English form of address *you* is translated into Polish in subtitles. The investigation is embedded in the framework of (im)politeness theory. The data used for the study comprise two types of non-professional renderings: amateur subtitles (fansubbing) and subtitles written by sophomore students of English Philology. The study proves that fansubbers' subtitles have a tendency to foreignise translation by ignoring the Polish sociocultural norms of terms of address, thus making it sound less polite and marked, while students' translations show a tendency to make the original dialogue more familiar to the target (secondary) audience, and make it more acceptable and polite.

Keywords

Polish forms of address, subtitles, corpus, non-professional translation, discursive approach, (im)politeness

Formy adresatywne i (nie)grzeczność: Analiza korpusowa polskich profesjonalnych i nieprofesjonalnych napisów filmowych

Abstrakt

Celem artykułu jest porównanie dwóch wersji tłumaczenia nieprofesjonalnego (napisów do filmów) pod względem stopnia ich (nie)grzeczności. Analizie poddano tłumaczenia amatorskie (tzw. fansubbings) oraz tłumaczenia studentów anglistyki. Badanie wskazuje, że tłumaczenia amatorskie brzmią mniej grzecznie niż przekłady studenckie. Różnica ta wynika m.in. z faktu, że studenci uwzględniają normy i konwencje grzeczności wyrażane przez formy adresatywne w języku polskim, podczas gdy amatorzy starają się, aby tłumaczenie pozostało bliskie oryginałowi, a tym samym przenoszą zwyczaje używania form adresatywnych z języka angielskiego (odmiany amerykańskiej) na grunt języka polskiego.

Słowa kluczowe

polskie formy adresatywne, napisy filmowe, korpus, tłumaczenie nieprofesjonalne, podejście dyskursywne, (nie)grzeczność

1. Introduction

The paper aims to analyse the occurrences of select polite and impolite forms of address in Polish subtitles written by two groups of non-professional translators: students and fansubbers. The two non-professional versions are also compared to the professional subtitles. As the study is embedded in the theoretical framework of im/politeness theories, some of the theoretical controversies and methodological problems are discussed in section 2. The paper deals with translations of forms of address from English into Polish; hence the Polish system of address terms is presented in section 3. The theoretical part is followed by section 4., which is entirely devoted to the presen-

tation of research aims and assumptions (section 4.1.), material and methods (section 4.2.), and the analysis of the data (section 4.3.). Discussion of the results is combined with conclusions in section 5.

2. (Im)politeness, discursive practices and address forms

The theory of politeness has gone a long way from universal and objective approaches to politeness viewed in the context of an ideal interlocutor (Brown and Levinson 1987 [1978]), with the early developments offered by Lakoff (1973) and Leech (1983), to subjective forms propounded much later by the discursive approach,¹ in particular by the internal (the user's) level of interaction analysis (Watts et al. 1992, Watts 2003, 2005, Mills 2003), where the interactants' perspective is central to analysis, to finally gravitate towards a more balanced approach, deployed primarily in *impoliteness* studies, which subsumes both the internal (the user's) and the external (the observer's) stands. The mixed approach is enacted, *inter alia*, by the socio-pragmatic approach (Culpeper 2011), which is akin to the observer's view, and the socio-interactional approach (Haugh 2015), which shows closer affinities with the user's perspective. Built on the premises of politeness studies and as a reaction to the view that incivility is only a failure in being polite, and thus it is mainly unintentional and accidental, impoliteness theory is marked for its intentionality,² internal diversity, and elusiveness of key

¹ One of the main claims of the discursive approach, yet not the only one, is that (im)politeness cannot be assumed to reside in linguistic forms. Watts (2002: 172), for example, stresses that "no linguistic expression can be automatically considered an example of politeness". This view is in tune with what was earlier expressed by Fraser and Nolen (1981: 96) who noted that "no sentence is inherently polite or impolite" and that "it is not expressions themselves but the conditions under which they are used that determines the judgement of politeness".

² Impoliteness was seen as intentional action in the early publications, mainly by Bousfield (2008). In more recent accounts, most scholars assume that impoliteness may be either intentional or non-intentional (e.g., Culpeper 2011, Terkourafi 2005).

terms and methodology. The multitude of theoretical approaches to verbal impoliteness makes it impossible to talk about a single impoliteness theory; any analysis follows a select school of research and the tenets thereof.

(Im)politeness is widely discussed in pragmatics, yet it is relatively rare and novel in the context of addresses.³ The discursive approach to (im)politeness has struck a chord with Clyne et al. (2009: 25), who draw on Watts (2003), and Vismans (2019), who relies in turn on Clyne et al. (2009). Both publications champion the presumption of discursive practices and rightly see the meaning of address terms emerging from meaning negotiation conducted by the interactants themselves through a “discursive struggle” (Watts 2003: 9) rather than from static, pre-existing semantic meaning. As Clyne et al. (2009: 25) put it: “address practices are relative and open to discursive negotiation”. In Clyne et al.’s (2009: 25) line of reasoning, the margin of negotiation is limited, as “individuals enter into any interaction with a set of at least partly shared assumptions about what is appropriate behaviour in the situation at hand, based on their knowledge about the world, their partly shared histories and cultural experiences, i.e., a (partly) shared background context”. These assumptions can be validated by “examining actual interactions” on the one hand, and on the other “by asking people about their experiences and views on address practices, as members of particular speech communities or social networks” (Clyne et al. 2009: 25). Suggestions like these are in general compatible with what is sanctioned by the methods of discursive analysis, in particular by Watts et al. (1992), Watts (2003) and Mills (2003); however, the methods are not unproblematic, as will be demonstrated below.

³ Whilst Zwicky (1974: 787) distinguished calls (whose aim is to catch attention, occurring in sentence-initial position) and addresses (whose position is other than sentence-initial) as two subtypes of vocatives, the terms addresses, terms or forms of address used here will encompass both calls and addresses in Zwicky’s sense. As Zwicky admits himself, the function of address is more general than calls and “all addresses are usable as calls” (1974: 791).

The discursive approach to (im)politeness (LPRG 2011), which represents the second wave in (im)politeness theories⁴ and is also known as impoliteness1, first-order politeness or emic (im)politeness,⁵ i.e., one which takes an internal view (the user's view), is promoted *inter alia* by Watts, Ide and Ehlich (1992), Eelen (2001: 252), Watts (2003, 2005), Locher (2004), Mills (2003, 2011), Locher and Watts (2005), and Geyer (2008). It criticises the etic view (the second-order politeness, impoliteness2, the external analyst's view), i.e., the metalinguistic evaluation conceived of by an external observer, as it "inevitably reproduces the researcher's own preconceptions" (Geyer 2008: 11). While the discursive approach has some undeniable contribution to the development of im/politeness studies, e.g., drawing researchers' attention to laymens' evaluations, focusing on long stretches of discourse that are above speech acts⁶ level, drifting away from universal, pre-conceptual norms and top-down analyses, centring on hearer's perspective, embedding it within the context of a Community of Practice (CoP)⁷, etc., it has some severe limitations, particularly as regards its methodology. Kádár and Haugh (2013: 40) voice a concern that the first-order impoliteness favours, in fact, the analytical perspective of a researcher while trying to conceptualise the im/politeness of

⁴ The first wave is the Brown and Levinson's (1987[1978]) theory of politeness as well as publications preceding it, i.e., Searle (1969), Lakoff (1973), Grice (1975), Leech (1983) (see Grainger 2011: 169–172, Culpeper 2011: 397, Bączkowska 2013).

⁵ These terms do not stand for exactly the same concepts, but their meanings considerably overlap. The terms emic and etic were first proposed by Pike (1954) and they originally come from the terms phonemic and phonetic. First- vs second-order politeness, in turn, are terms originally used by Watts, Ide and Ehlich (1992) and popularised by Eelen (2001), Watts (2003) and Grainger (2011).

⁶ Werkhofer (2005: 171) even holds that politeness may be built up over several turns and thus it runs counter to what he calls the "mentalistic approach" focusing on what the hearer is thinking about (the "within the head" approach).

⁷ This term was proposed by Wenger (1988) who defines Community of Practice as a group of people with mutual engagement (doing things together) in a joint enterprise (a negotiated action with mutual accountability) who share repertoire (concepts, tools, styles, historical events, etc.) (Wegner 1988: 73). In earlier studies, the term *speech community* was preferred (see e.g., Lyons 1970, Labov 1972).

the lay interactants. Thus, the first-order impoliteness (i.e., the Wattsian discursive approach) is criticised for precisely the same problem as the second-order impoliteness. In order to understand the problematic methodology of the first-order impoliteness, which started the so-called discursive turn, it will now be elaborated in more detail.

In essence, the “discursive turn” (Kádár and Haugh 2013: 6, van der Bom and Mills 2015: 181) assumes that the (negotiated) meanings are affordable to the interactants (laypersons) but not to the observers. Thus, an analyst (researcher) has no or, at best, minute access to the interlocutors’ understanding, perception and experiences of the negotiated meaning (Watts et al. 1992, Watts 2005, Locher and Watts 2005, Mills 2003). Admittedly, the situated interaction, i.e., ongoing in real time and seen from the participant’s perspective, is the only legitimate source of information for a study of interaction (for criticism of this view, see Terkourafi 2005: 241). The cornerstone of the discursive approach is ventured by Mills as follows: “the focus is on what the language used means to the participants, including both speaker and hearer, whether the participants themselves classify the utterances as polite or impolite, how they come to make those judgements, and what information and cues inform those decisions about whether someone has been polite and impolite.” (Mills 2003: 5). The discursive approach draws on the claim that im/politeness “emerges at a discourse level, over stretches of talk and across communities of speakers and hearers” (Mills 2003: 70). It arises out of instances of interaction, and thus, it is dynamic, praxis-embedded, and open to adaptation with a particular group (Watts 2005: xviii). It is interactant-informed and individual-oriented, and it instantiates social interaction, where social or CoP norms play a role. As a result, they are post-modern, as opposed to the so-called modernist take that abstracts interaction away from its participants (Watts 2005: xlii). The interactants’ perspective (the lay concept) is of vital importance, and Watts (2005: xxi) argues that it is the laypersons’ conceptualisations which should be the bedrock of a postmodernist approach (rather than the theoretical con-

structs of some observer-expert). Thus, the aim of politeness research should be to “locate possible realisations of polite or impolite behavior” on the one hand and to assess “how the members themselves may have evaluated that behavior” (Watts 2003: 19–20) on the other. Accordingly, it is not the linguistic expressions chosen by participants that a researcher should focus on but the process of arriving at participants’ evaluations of the language utilised in interaction (see Kecskes 2017: 11). As can be seen, (im)politeness theory has shifted from pragmatic analyses (as practised by the first wave of politeness studies, represented by *inter alia* Brown and Levinson (1987[1978]) and Leech (1983), to a social stand promoted by the discursive approach, and from a theory-driven to data-driven methodology.

Mills (2003) takes a less firm stand on discursive practices than Watts as, while she does emphasise on many occasions that participants are the source of the emerging meaning and not the researcher, and thus ascribes the notion ‘discursive’ to (im)politeness¹, she does not seem to entirely disallow the external perspective inherent in (im)politeness²: “discursive approaches *tend* to focus on first-order evaluations (Mills 2017: 15, emphasis mine), which suggests that for Mills the notion discursive, while primarily emic, is also to some extent etic. Mills (2003: 82) resorts to the idea of meta-discourse levels of analysis proposed by Taylor (1992), i.e., the intellectual meta-discourse and the practical meta-discourse. Mills holds that unlike the intellectual meta-discourse, which is the “theoretical analysis and thought”, the practical meta-discourse applies to the discursive approach, which is “the thinking about what has been said previously in terms of the impact it makes on our relationships.” Thus, some meta-analysis is permitted by Mills (2003), which is still conducted at the level of participants’ conceptualisations but is realised through *post-hoc* reporting and evaluations. The practical meta-discourse is not an internal perspective through and through, but it is not an external one either, as the latter is typically associated with the non-participant’s perspective.

Furthermore, in the introduction to the book devoted to discursive approaches to politeness (LPRG 2011: 5), she makes a disclaimer in the footnote from which it transpires that the book couches studies which inscribe in the discursive approach to varying degrees, that the discursive approach is methodologically heterogeneous, and “a wide range of approaches can come within its ambit”. The term discursive approach should thus “be seen as an umbrella term for a fusion of theoretical and methodological strands” (van der Bom and Mills 2015: 181). By so doing, she actually opens the possibility of an extended understanding of the term discursive, originally ascribed to (im)politeness¹ only, and encourages other approaches diverting from impoliteness¹, possibly also those encompassing impoliteness², at least to some degree. In her later publication, Mills (2017: 17) expounds her “modified discursive approach”, which she dubs the discursive-materialistic approach⁸. Mills (2017: 17) maintains that this new version belongs to the third-wave approach (the emic-etic one). Haugh and Culpeper (2018) also include this modified version in the third-wave approaches, yet they see it as the closest one to the user’s perspective (i.e., strongly emic-oriented). It seems that the originally somewhat radical stand represented by the discursive approach (particularly epitomised by Watts’ argumentation) has evolved and relaxed conceptually. However, as shown below, it still does not cope well with adjusting the more flexible definitional framework to the invariably rigid methodological requirements.

The initial popularity garnered by the discursive approach was dented by its critique, particularly by the new approaches to im/politeness collectively called the third wave of im/politeness research (Terkourafi 2005, Locher 2006, Spencer-Oatey 2008, Garcés-Conejos Blitvich 2010, Kádár and Haugh 2013). These are approaches to im/politeness, which assume an intersection of the internal (emic) and external (etic, impoliteness² or second-order politeness) perspectives, i.e., the interactants’

⁸ In this version, she takes into account ideological and class-based divisions and analyses how they influence the interaction style of an individual (Mills 2017: 19-20).

and the researcher's take. The mid-way stand, dubbed 'interactional approach' by Grainger (2011), 'integrative pragmatics' by Culpeper and Haugh (2014: 266) or 'pragmatic-discursive approach' by Félix-Brasdefer (2015), is represented by a number of (im)politeness researchers, such as Haugh (2007), Culpeper (2011), Grainger (2011), Kádár and Haugh (2013), Culpeper and Haugh (2014), Félix-Brasdefer (2015). The perspective combining the emic (the user-participant) core approach, dubbed discursive by Watts and Mills *sensu stricto*, with the etic (the observer-analyst) perspective considerably drifts away from the original concept of discursiveness as proposed in particular by Watts et al. (1992), Watts (2005), and to some extent also by Mills (2003, for discussion, see below).

The original term discursive approach, rapidly gaining popularity, is, all too often, overused and overinterpreted or even misinterpreted (van der Bom and Mills 2015: 180). It relies on the internal real-time evaluations by interactants themselves rather than *post-factum* measures. The definition of the discursive approach is stated by van der Bom and Mills (2015: 181): "the discursive approach is not simply a critique, but constitutes a mode of analysis itself"; thus, it should necessarily entail its methodology of the inner perspective. Hence audio-recording, interviews, role plays, focus groups and questionnaires for participants are adopted to elicit the inner perspective (Mills 2003: 45, van der Bom and Mills 2015: 188). However, for these elicitation methods, Mills was criticised by Terkourafi (2005). According to Terkourafi, administering questionnaires and interviews is, in fact, resorting to *post-factum* measures. Kádár (2013: 117) also expresses critical thoughts on such elicitation methods. He claims that the subjective perspective reported by an interactant involves a retrospective narrative and thus is no longer a direct account of the inner perspective but of a story seen as if from the outside. Kádár (2013: 171) notes that "even those who originally participated in an interaction become semi-observers when they narrate and reflect on an event retrospectively". Moreover, when they report on an interaction, they may intentionally bias their account, making it more euphemistic if they

are aware that their abusive verbal behaviour is morally unacceptable. This is also frequently the case of other participants' accounts, who do not take active participation in an interaction, the so-called "onlookers" (Kádár 2013: 172). The *post-hoc* methods employed to elicit the participants' views thus do not seem to do justice to the first-order impoliteness approach. In fact, Mills (2003: 45) herself admits that in the case of *post-hoc* reporting, "it is getting no nearer in essence to what really went on, as it is simply another text, (...) only this time with the analyst". Put differently, the participant also has the status of the analyst for Mills. As most scholars (for details see, e.g., Kádár and Haugh 2013: 87) identify the analyst with the external perspective associated with second-order impoliteness (the researcher or a non-participating analyst), rather than the (meta-) participant's retrospective accounts, the mixed (third-wave) approaches that permit internal and external assessments are not aligned with the discursive approach built on impoliteness¹ (and analyst-participant) only. In fact, they are often critical of it (cf. e.g., Terkourafi's frame-based approach 2005, Culpeper's socio-pragmatic approach 2011, Haugh's interactional approach 2014, etc.). Mill's 'analyst', who resorts to (Taylor's) 'practical meta-discourse' while reporting on an interaction s/he participated in, is an external participant (called *etic* by Kádár and Haugh, 2013: 87) but only on the first-order impoliteness level as s/he has not the status of an external observer (like the research or even the layperson do).

Given the theory of discursive practices, as promoted by *inter alia* Watts et al. (1992), Locher (2004), Watts (2005), and Mills (2003), and its convincing criticism, a description of any film discourse translation in relation to (im)politeness is never an instantiation of solely first-order impoliteness analysis. Firstly, the researcher describing a translated version focuses on the decisions taken by the translators. The rendition of the original dialogue lists already forms the external analyst's perspective. Additionally, the researcher's perspective is mapped onto the translator's perspective, and thus the external perspective occurs twice inasmuch as both measures are *post-factum*, meta-

linguistic, and filtered through the interpretations of the analyst-translator and analyst-researcher. Secondly, what the analyst-translator and analyst-researcher have access to is only the actors' lines, which are not naturally occurring data but a constructed dialogue, an "artistic verisimilitude" (Lakoff and Tannen 1979: 581). Therefore, one cannot talk about genuine discursive struggle or (im)politeness negotiation. All the lines actors utter are previously written specifically for the purpose of a film by (a) scriptwriter(s), then learnt by heart by actors, and next filmed, often after a number of rehearsals and trials. The whole process is pre-planned and highly controlled and the language is often multiple-authored, less spontaneous than in natural speech and simplified, mainly prosodically and syntactically (Bednarek 2019, Bączkowska 2022b). Even if we consider film discourse to be a convincing imitation of natural discourse, and expand the notion of the emic level to encompass it, the analysis can, at best, be emic-etic.

Considering the definition of the notion *emic* originally proposed by Pike (1954) and further expounded especially by the third wave of (im)politeness theorists, to which this paper adheres, particularly by Kádár and Haugh (2013: 94–96), some degree of the emic element is also present in this analysis. In line with these authors, the emic level is understood as underlying expectancies regarding evaluations of what is conceptualised⁹ as socially (in)appropriate by the insiders across a CoP, which stem from institutionalised norms (see also Eelen 2001: 76). As a result, the analysis proposed in this paper is primarily an etic one, certainly on the methodological level, yet, on the emic level, it makes reference to social and cultural conventions relevant to the system of behaviour of insiders, which define (im)polite behaviour and allow participants (and viewers) to

⁹ The user conceptualisation of social norms of moral behaviour forms the first-order (im)politeness perspective, while their interpretation (i.e. their understanding based on actual realisations in an interaction) by an observer belongs to the second-order (im)politeness perspective (Kádár and Haugh 2013: 104).

build some expectancies of moral verbal behaviour (see Kádár and Haugh 2013: 94–96).

On a more practical note, politeness can be expressed by choosing relevant addresses or through lexical choices that are evaluated as polite in a particular language or a speech community, i.e., ones that encode social distance and respect (titles, formal forms). Along with these, diminutives and hypocoristic forms may also encode politeness if utilised among interactants already on informal terms. On the other hand, impoliteness or the wish to offend a target may be signalled by resorting to terms of address indicative of close relations when they address interactants one has only formal relationships with, i.e., by breaching social distance, or, even worse, by deploying offensive address terms, especially those deriving from pejorative and emotion-laden nouns or adjectives (e.g., *thicko*, *fatty*, *pig*, etc.).

3. Forms of address in Polish

The widely known and often criticised (e.g., Slobin 1963; Stone 1977: 491–493; Braun 1988; Hill 2014) division of languages into those using the formal (*V*, from Latin *vos*) versus the informal (*T*, from Latin *tu*) addresses proposed by Brown and Gilman (1960), applies to Polish only partially (Bączkowska 2019, Bączkowska 2022a). As will be shown below, in Polish, the forms of address comprise not only *V*, *T*, but also *P* (*pan*), and even occasionally *O* (*oni*) (Bączkowska 2019 and references therein).

Tu (T). The informal *T* address may be realised by the Polish *ty* (“you”), yet other forms are equally frequent (or even more frequent) as Polish is a pro-drop language. For example, verbal addresses (Rusiecki 2008) are allowed wherein verb endings coding the 2nd person singular take over the function of *T* marking (mostly *-esz*, e.g., *piszesz* meaning “you write/are writing”, example 1). Another option is resorting to verbless constructions (example 2). The use of *ty* (“you”) is, in fact, burdened with some axiological charge, either positive (example 3) or negative (exam-

ple 4, 5), or is typical of slang (example 5), and then it remains relatively neutral.

- (1) *Co robisz?* (“What are **you** doing?”)
- (2) *Ty(Pani) z Polski?* (“You [+informal/(formal), +sing.] Ø from Poland?”)
- (3) *Nie jestem przystojny, ale ty jesteś.* [“I’m not handsome but you are”]
- (4) *Ej, ty, uważaj co mówisz.* [“Hey, you, watch out what you are saying”]
- (5) *Na litość boską, co ty tam robisz?* (“For God’s sake, what are you doing there?”)
- (6) *Ty, masz fajki?* (“You [+voc.], do you have ciggies?”)

Vos (V). Whilst the *V* option is practically unused in contemporary Polish, it must be mentioned that it was present in the past, specifically from the middle of the 18th century till the first world war, and later when it was unsuccessfully reintroduced in the communist times in official language following the words *obywatelu* (“citizen” in the vocative) and *towarzyszu* (“comrade” in the vocative) (example 7; Stone 1977: 493, Łaziński 2006: 441–3, Huszcza 2006: 11). There are also contexts wherein the official *V* equivalent is intertwined in a sentence with unofficial 3rd person plural (example 8) or, less seldom, with 3rd person singular (example 9). Both cases create a sense of strong impoliteness encoded by the singular form. The contexts deploying *V* sound old-fashioned, and, as mentioned above, they are no longer, or very rarely, used in contemporary Polish. They are, however, common in some other Slavic languages, such as Russian or Bulgarian (Sosnowski 2015: 323).

- (7) *Towarzyszu, dokąd idzie**cie**?* (“Comrade, where are **you** [2nd pers. pl.] going?”)
- (8) *Czy wy**ście** to napis**ali**?* (“Did you [2nd pers. pl.] write [3rd pers. pl.] it?”) (after Stone 1977: 493)
- (9) *Czy wy**ście** to napis**ał**?* (“Did you [2nd pers. pl.] write [2nd pers. sing.] it?”) (after Stone 1977: 493)

Both *V* addresses discussed above and *O* addresses mentioned below substantiate *pluralis maiestaticus* (Sikora 1993), wherein an addressee is invoked by words in plural, either in the 2nd person (*V*) or 3rd person (*O*).

Oni (O). The form that Stone (1977) calls *O* (from Polish *oni* meaning “they”) encodes *oni/one* (“they” + masculine and “they” + feminine), i.e., the 3rd person plural when addressing 2nd person singular. The *O* form is not widely used today in Polish (Zaręba 1974: 378–388), except for some humorous contexts (especially when talking to children, example 10). The *O* addresses have been noted in Czech (Kretzenbacher et al. 2013), where it is still used today (originally a calque of the German *Sie*).

(10) *Gdzie są moje małe córeczki?* [“Where are [3rd pers. pl.] my little daughters [+plural, +diminutive]?”]

Pan (P). The form *P* stands for *pan* (masculine singular, example 11), *pani* (feminine singular, example 12) or *Państwo* (plural), which is a formal address. Originally a noun meaning a Polish landowner, landlord, or a superior person (Klemensiewicz 1946: 34–35), which appeared already in the 13th century (Łaziński 2006: 23), *pan* has changed its status into a pronoun (Klemensiewicz 1946, Pisarkowa 1979: 6–7, Stone 1981, Sikora 1993: 300, Huszcza 2006: 97). The form *pan* replaced the *V* form, which started to be phased out by *pan* in the language of the gentry in the second half of the 16th and the 17th century and became popular in urban areas in the 19th century. The *V* address was gradually ousted and marginalised, frequently utilised only in contacts between the middle class and people of lower social status, typically addressing servants (Sikora 1993). This shift lasted until 1945, yet in a rural context, it was still noted in the ‘70s (at least in the south of Poland), where it was a sign of respect towards the addressee (Sikora 1993: 299–303). The *P* form is typically deployed in Polish with the first name (example 12), and it is considered impolite when patterns with

the surname, for example, *Panie Kowalski* (example 13; Miodek 1991). Interestingly, *pan/i* can also be intertwined with the first name, which in Polish marks informal/intimate relationships or shortening social distance (example 14) with polite overtones or with the verb ending of 2nd person singular to mark informality (Sikora 1993: 300) and an impolite tone of disrespect (Łaziński 2006) typical of colloquial Polish (Sikora 1993: 300) (example 15). Finally, *pan/i* may co-occur with *proszę* (“please”) to form the phrase *proszę pana/pani*, which is a polite and formal form of address that disallows co-occurrence with first or last name and thus makes it sound less personal (example 16).

- (11) *Gdzie pan mieszka?* [“Where do you [+formal, +singular] live [3rd pers. sing.]?”]
 (12) *Pani Mario, proszę do mnie jutro zadzwonić.* (“Marry [+formal, +sing.], please call me tomorrow.”)
 (13) *Pani Kowalska, proszę do mnie jutro zadzwonić.* (“Mrs Kowalska [+formal, +sing.], please call me tomorrow.”)
 (14) *Marysiu, zrobi mi pani kawę?* (“Marry [+diminutive], will you [+formal] make coffee for me?”)
 (15) *Masz pan zapalki?* (“Do you [+formal, +sing.] have [+2nd pers. sing, +informal] matches?”)
 (16) *Proszę pani, proszę do mnie zadzwonić jutro.* (“(Dear) madam [+polite], please call me tomorrow.”)

Taken together, the Polish terms of address, theoretically, may take one of the four forms: *V*, *T*, *P*, *O*. However, the *V* form is old fashioned, and the *O* form is uncommon; thus, the most typical forms of address in Polish are *T* for informal terms and *P* for formal.

4. Research assumptions and results

4.1. Preliminary assumptions and research aims

The addresses presented here were analysed only in the context of dialogues (dilogues or polylogues, Kerbrat-Orechioni 2004) when address terms were directed towards ratified (Goffman

1981) addressees; thus, addresses directed to overhearers and eavesdroppers were ignored. Furthermore, the analysis describes translators' choices; this necessarily entails the researcher's interpretation of translators' interpretations. Accordingly, the study primarily investigates (im)politeness², viz., the external perspective. The analysis at both levels (the researcher's and the translators') is *post-factum* and relies, on the one hand, on some theoretical frameworks of how (im)politeness should be expressed in Polish, allowing for the Polish-specific linguistic and sociocultural (im/politeness) norms, and on the other, on how translation should proceed (in particular which translation and/or subtitling strategies/techniques to employ).

The study presented in what follows revolves around several issues regarding the translation of various forms of address across English and Polish, which are embedded in the more general aspect of (im)politeness. The main research question is which subtitling version, the amateurs' or the students', is more polite and which is more impolite. This will be checked by making reference to the conventions of Polish addresses (discussed in Section 3). The two non-professional renditions will also be compared with the Polish professional subtitling version to answer the question which non-professional version is closer to the professional one. Since the professional version will be a reference point, it will also be checked whether it is marked for politeness embedded in the Polish culture- and language-specificity. Some references will occasionally be made to the original version to determine whether any cultural or linguistic transfer from the film's original language can be observed. Addresses in the context of (im)politeness will be analysed on the basis of (i) the Polish form of polite address *proszę pana* ("Sir") (section 4.1.1), (ii) the use of marked *ty* ("you") encoding a positive or a negative axiological charge (section 4.1.2), (iii) the expressions of endearment (4.1.3), and (iv) the deployment of the formal address *pan/i* vs the informal first name (FN), and the formal *pan/i* vs the formal first name plus last name (FN+LN) (section 4.1.4).

4.2. Material and methods

The corpus the study is based on is ca. 25 thousand words in size, and it consists of two subcorpora: non-professional subtitles produced by sophomore students of English philology (2nd year students of MA studies), and amateur subtitles retrieved from the web (www.napisy.info). Both translations are non-professional, yet the students have a wide background in cultural studies, linguistics and translation theory, which they receive in the course of their university studies, whereas amateurs usually do not have such a theoretical background but often have a great deal of practical experience in film translation. The students' translations constitute a part of the Learner Corpus of Subtitles project (the LeCoS project) conducted between the years 2010 and 2015, which overall involved over one hundred students and is ca. 100,000 tokens in size (Bączkowska 2015a, 2015b, 2016, 2019, 2022).

On the other hand, the amateur translations are written by fansubbers, most probably by the HATAK group, presumably the oldest and best-known Polish fansubbing group. Fansubbing is a consumer-driven activity of subtitling by fans for other fans. Initially, it arose in the Japanese anime subculture in the US, where anime films were considered inappropriate content-wise (containing sexual and violent overtones) and ultimately banned from the US film market. Consequently, informal communities of amateur subtitlers started to be formed in the '80s (O'Hagan 2009). Fansubbing is usually conducted by fan-based communities consisting of people, either loosely gathered with *ad-hoc* norms or well-defined groups with workflow, internal protocols and subtitling guidelines, who are usually without formal training in subtitling, work without pecuniary remuneration using unauthorised copies of films and, more often, TV series, and frequently breaching authorship rights (see Massida 2020 for more details).

The HATAK group declares on their website (grupahatak.pl) that they provide translations where, unlike in the case of voice-over, nothing important is omitted in their subtitles. This appro-

ach contradicts the general rules of subtitling, particularly the prescription of text reduction by around 30 % – 40 % in subtitles (Tomaszkiewicz 2006: 113). The realisation of the assumption expressed by the HATAK group on their website is visible in the data at hand, wherein the reduction of the original text is relatively small and amounts to around 14 %, which compared to the students' versions (ca. 30 %), and the professional subtitler (ca. 40 %) is relatively low.

The LeCoS (Le) subcorpus used for this analysis consists of three independent translations of the film *What Women Want* (aired in 2000, directed by Nancy Meyer) and three amateurs' renderings (Fansubs) downloaded from the web. The students had a training session on how to subtitle in line with the generally accepted subtitling rules (e.g., considering the so-called time and space constraints), along with a series of lectures on translation and subtitling strategies. They used the free online Subtitle Workshop software (v. 6.0; <http://subworkshop.sourceforge.net/>) to write their subtitles, which was installed prior to the subtitling sessions in the university computer lab. At the time of the corpus collection, only three fansubbers' versions of this particular film were available, so to balance the corpus, three students' translations were randomly selected for the present study. Altogether, six full versions of the subtitles were collected. The non-professional versions, the professional translation retrieved from the DVD version, and the original dialogues were all stored on the Sketch Engine (<https://www.sketchengine.eu>), a commercial corpus management system. All data were retrieved from the corpora automatically with the use of the CQL¹⁰ that allows batch retrieval, i.e., categories of data, such as all instances of verbs in the second person singular or all forms of the *P* address (regarding declinations, plurality, gender).

The lexical level of analysis of the translations is enriched by the examination of contexts, both at the speech act level and

¹⁰ Corpus Query Language is partially based on regex and allows extracting classes of words or phrases based on, *inter alia*, PoS tags (for more details, see Bączkowska 2020).

beyond it, to secure precision in the study at hand. Thus, the investigation presented here subscribes to the context-dependent evaluations of translations. Each context containing a form of address was copied to an Excel spreadsheet and tagged for endearment (politeness), positive vs negative marked *ty* (politeness, impoliteness), and for the use of *pan/i*. Larger contexts were found in the film and carefully rehearsed to gain further contextual clues, including nonverbal ones.

4.3. Research results

4.3.1. *Proszę pana* and P+LN

In Polish, politeness can be achieved by leveraging the address *proszę pana* (lit. “I please you, sir”). In the LeCoS data, this address has 9 occurrences (1 in Le1¹¹, 5 in Le2 and 3 in Le3), while in the fansubbers’ 3, and it was only applied by one person (Fa2) (Figure 1). This speaks for more polite forms in the students’ translation in our data than the amateurs’ subtitles under study.

,580 --> 00:07:41,875 Tak. Dziękuję Ci! 115 00:07:41,917 --> 00:07:44,628 *Proszę Pana* , to było inspirujące 116 00:07:44,712 --> 00:07:46,630 Wiem 117 00:07:48, :30,817 - Ale pluca... 74 00:05:30,900 --> 00:05:34,738 - Wspaniałego dnia, *proszę Pana* , - Wzajemnie. 75 00:05:37,365 --> 00:05:40,202 Spocznij. 76 00:05:54,467 --> 00:07:41,875 To randka. Dziękuję. 115 00:07:41,917 --> 00:07:44,628 *Proszę Pana* , To było inspirujące. 116 00:07:44,712 --> 00:07:46,630 Wiem. 117 00:07:48 31:21,371 Ja? Nic. 488 00:31:21,455 --> 00:31:25,709 - Jesteś pewna? - Tak, *proszę Pana* . 489 00:31:29,297 --> 00:31:31,215 Wiesz co? Chyba się dziaś!aj przejdę. 490 --> 00:31:34,552 - Trochę świeżego powietrza dobrze mi zrobił. - Milego dnia, *proszę Pana* ... 491 00:31:34,636 --> 00:31:38,598 Ze swoim świetnym tyłeczkiem 05:30,817 - Niezły gwizdek. 74 00:05:30,900 --> 00:05:34,738 - Milego dnia, *proszę pana* , - Nawzajem.. 75 00:05:37,365 --> 00:05:40,202 Spokojnie . 76 00:05:54,467 --> 00:07:41,875 Tak, randka. Dzięk! 115 00:07:41,917 --> 00:07:44,628 *Proszę pana* , to było świetne. 116 00:07:44,712 --> 00:07:46,630 Wiem . 117 00:07:48,341 620 Dzień dobry dziewczęta. 155 00:09:32,704 --> 00:09:34,914 Dzień dobry, *proszę pana* . 156 00:09:34,998 --> 00:09:37,959 Hej, czy ktoś wie dlaczego zebranie :31,299 --> 00:31:34,552 - Potrzebuję świeżego powietrza. - Udanego dnia, *proszę pana* 491 00:31:34,636 --> 00:31:38,598 Z takim tyłeczkiem ! 492 00:31:38,682 -->

Figure 1

Concordances with the Polish formal address
proszę pana in students’ subtitles

In the fansubbing corpus, an interesting example of the use of *pan* (“sir”) is illustrated by example 17, wherein the formality of address is encoded twice: first by the address *pan* (+genitive, +possessive) and then by *pan* +LN (+vocative). Other renditions involved either the possessive form of *pan* or the vocative form (in the nominative case). Using CQL formula (with the distance

¹¹ “Le” stands for learner’s translation while “Fa” for fansubbing rendering.

between *pan* and LN up to 20 words in between), however, allowed me to tease out only one example with double formality encoding in the Fansubbing corpus as well as in the LeCoS data and the professional subtitles. Strong formality and politeness are thus uncommon in the data at hand. As noticed by Bączkowska (2022: 86), however, the address consisting of *pan*+LN (without the double repetition of *pan* as mentioned above) is relatively common in both non-professional versions (with a slightly higher result for fansubbers).

- (17) **Pańskie** [+3rd per. sg., +formal, +possessive] włosy wyglądają dzisiaj naprawdę dobrze, **panie Marshall** [+3rd per. sg. + LN, +formal]. [Fa2-00:35:03] (“Your hair looks really nice today, Mr Marshall”)

4.3.2. Marked *ty*

As Polish is a pro-drop language, the pronoun in the nominative is habitually omitted. When it is purposefully used, it has an emphatic function. The number of occurrences of the pronoun *ty* (“you”) in the nominative is illustrated in Figure 2.

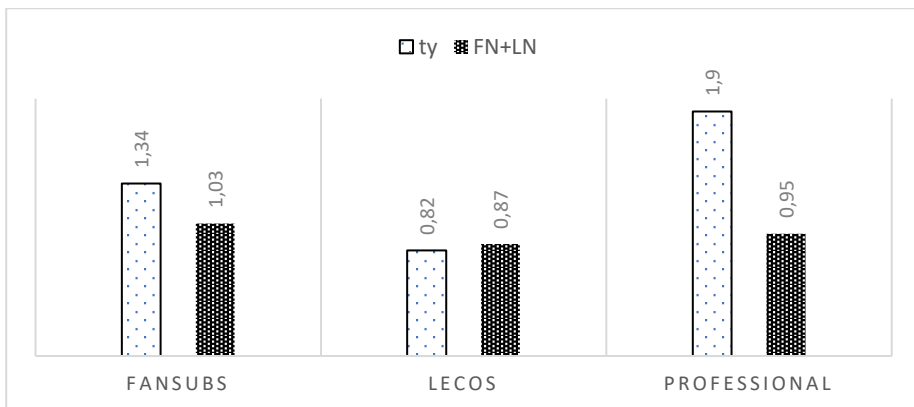


Figure 2

The incidence of *ty* versus FN+LN against all words in the three subcorpora

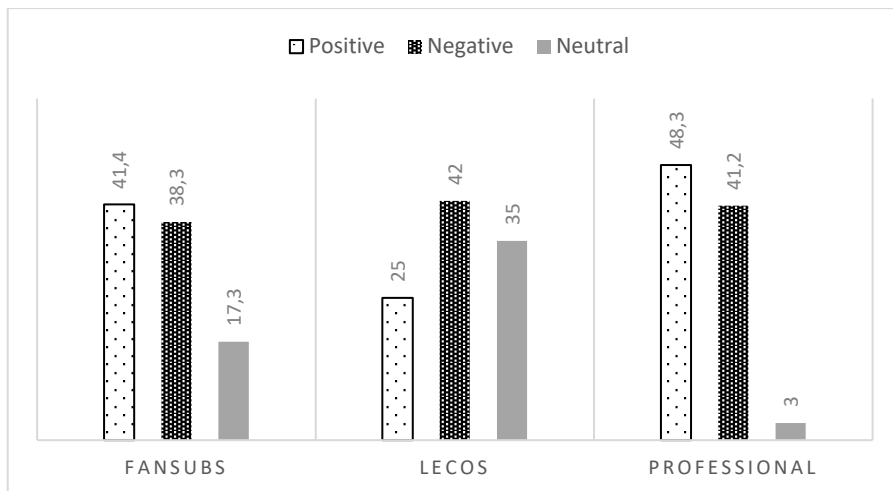


Figure 3

The axiological charge of *ty*

Marked *ty* is deployed much more often by amateur subtitlers than by student subtitlers. Students tend to omit *ty* altogether or to resort to FN+LN. Almost all the instances in the learner corpus contain this form of address with the capital letter (*Ty*), which is a very polite form. Interestingly, students' renditions of *you* into marked *ty/Ty* appear primarily in negative contexts to criticise or express irony (Figure 3). Impoliteness is thus more often expressed by dint of marked *ty* in LeCoS data and professional translation. On the other hand, fansubbers and the professional translator resorted to *ty* also (in fact, most often) to encode positive contexts, and, in this respect, the amateurs' rendering bears more similarity to the professional translation. A Chi-square test shows that the differences among the two non-professional and the professional subtitles in terms of axiological charge are statistically significant ($\chi^2(2, N = 159) = 11.28, p < .05$).

It can be observed that marked *ty* occurs more seldom in the non-professional translations where the number of first name addresses (mostly *Nick* and *Darcy*, the two main characters) is high. The data were also searched for contexts where *ty* and *Nick* occurred in one utterance, yet no instances were found.

This lends support to the tentative observation that whenever there was emphatic *ty* occurring in the nominative, the first name was not present in the same utterance. Some examples from the fansubbers' corpus are shown below (examples 18–21).

- (18) *I co ty wiesz o związkach?* (“And what do **you** know about relationships?”) [Fa1-01:20:03]
 (19) *Co ty robisz? Musimy iść.* (“What are **you** doing? We have to go.”) [Fa1-00:34:34]
 (20) *Ty mówisz po wenusjańsku* (“**You** speak Venusian”) [Fa1-00:49:48]
 (21) *Chcę, żebyś ty to wziął.* (“I want **you** to take it”) [Fa1-01:39:57]

4.3.3. Endearment – hypocoristic forms

Endearments are types of forms of address that “mark a bond of closeness and affection” (Biber et al. 1999: 1110). They play a unique role in maintaining and reinforcing relationships, “usually marking a bond of closeness and affection between close family members, sexual partners, and other ‘favourite’ people” (Biber et al. 1999: 1110).

In the original dialogue, one endearment addressee is used, which is *honey* (11 times, once also *hon*). The address *sweetie* also appears but only once, and it is reserved for the context with Nick, the father, talking to his teenage daughter (Alex). On the other hand, *honey* is employed more often with adult addressees, mostly Nick’s lover Lola, Nick’s cleaning lady talking (thinking, in fact) to herself addressing Nick, but also when Nick or his ex-wife (Gigi) addresses Alex. Interestingly enough, in the Polish amateur subtitles, two address terms are observable, *skarbie* (“baby”, lit. “treasure”, +voc., 23 times) and *kochanie* (“honey”, +voc., 11 times), which occur equally often. *Skarbie* is essentially present when Nick talks to Lola, while *kochanie* is reserved for all the other contexts (talks between adults, a talk between Nick and his daughter Alex, and between Alex and her mother, Gigi). In the students’ translations, on the other hand, there are six options: *laleczko* (“dolly” +dim., +voc., once), *śłonko*

(“sun” +dim., +voc., twice), *skarbie* (“lit. treasure”,+voc., 20 times), and *kochanie* (“honey” +voc., 28 times), *moja najdroższa* (“my dearest + voc.”, once), *różyczko* (“rose” +dim., +voc., once). Even when Nick makes a jibe at Alex about her being slightly chubby by calling her *pumpkin*, non-professional translators rendered it as a polite hypocoristic form of address, except for one student who resorted to *pulpet* [+voc.] (‘meatball’), referring in this way to the roundish shape of the teenage (15-year-old) girl. There is a wider variety of hypocoristic forms in the students' translations. The two most common translations of *honey* and *baby* are “kochanie” and “skarbie”.

Both types of non-professional translation betray symptoms of overusing hypocoristic forms relative to the professional version, with students' translations being on the top. Thus, in terms of quality (i.e., creative variants) and quantity (Figure 4), the students' translations demonstrate the most intense use of hypocoristic terms of address.

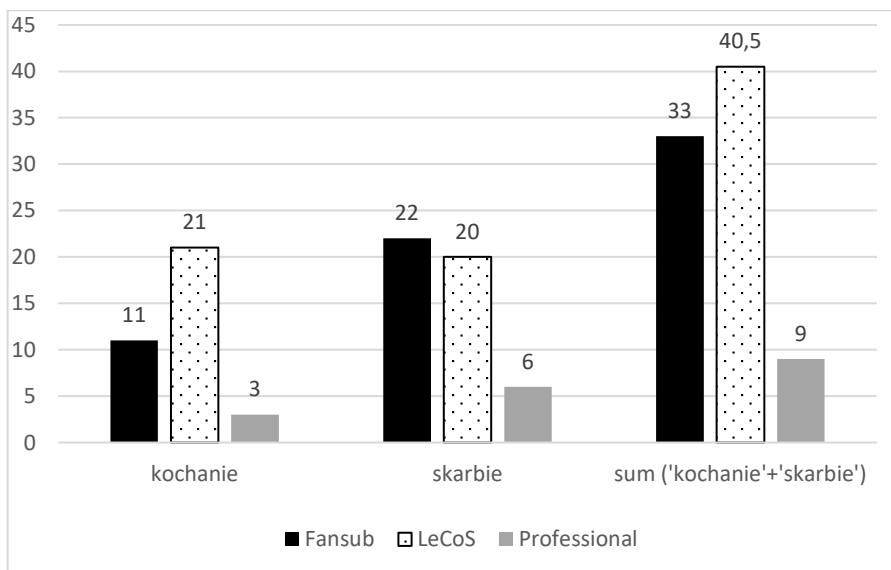


Figure 4

The incidence of the most frequent hypocoristic forms

4.3.4. *Pan/i* vs FL and *pan/i* vs FL+LN

Comparing the occurrence of FN against *pan/i* (Figure 5), it must be noted that the highest incidence of FN is observed in the fansubbing versions. Further to this, while in the students' and professional translations the number of occurrences of FN and *pan/i* is comparable, fansubbers preferred the informal forms of address encoded by first names at the cost of the formal *pan/i*. The tendency of fansubbers towards a more informal style in translation is clearly noticeable. Moreover, the occurrences of *pan/i* versus first names followed by the last name (e.g., *Nick Marshall*) demonstrate that fansubbers opted for the formal address (FL+LN), which sounds awkward in Polish and is a direct translation from American English. Students (as well as the professional translator) went for a more acceptable in Polish and a more polite form *pan/i* (Figure 6). Overall, the amateur version is the least formal of the three translations, with the formal form *pan/i* occurring least often and FN occurring predominantly; the students' version is rather formal, with a high incidence of *pan/i* and a low incidence of first names; professional translation is also formal (high *pan/i* occurrences) and is marked for a high use of names (including *pan* + LN).

A few words are in order regarding the rendering of FL+LN into Polish. While in naturally occurring encounters with participants speaking Polish, the *pan* + LN is perceived as somewhat abrasive and even disparaging (e.g., *panie Kowalski*, Eng.: *Mr Kowalski*), in film translation, it does not strike so much as impolite. The reason might reside in the fact that the more polite version thereof would be a combination of *pan* + FN, that is the use of the first name instead of the last name (e.g., *panie Piotrze*, Eng. ? *Mr Peter*), but foreign first names are resistant to vocative cases in Polish (? *panie Nicku*) and with the Polish inflections responsible for the vocative case (-*u*) they sound both funny and untypical. They also sound unnatural and could be perceived as impolite when used in the nominative, i.e., without the Polish inflections (e.g., *pani Kate* or ? *panie Nick*). While deciding what is and what is not a polite form of address and how to translate

addresses, it is thus essential to map them onto the sociocultural expectations and language conventions of the target language. Given these arguments, the Polish translation of *Mr Marshall* as *Panie Marshall* is probably the best option.

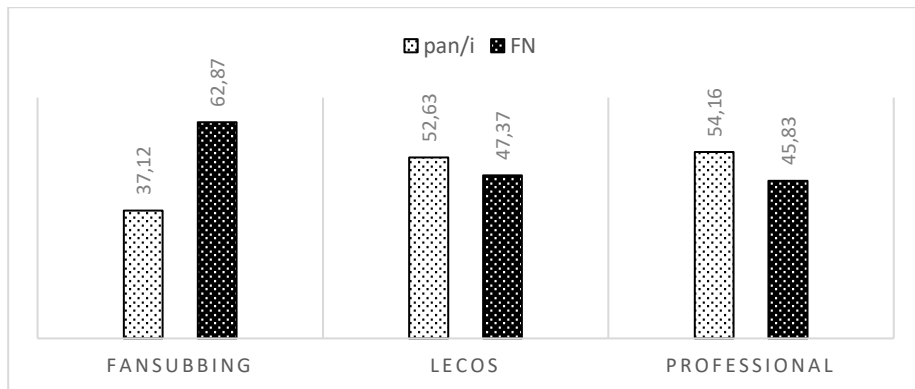


Figure 5

Corrected¹² incidence of *pan/i* versus first names across three subcorpora in percentages

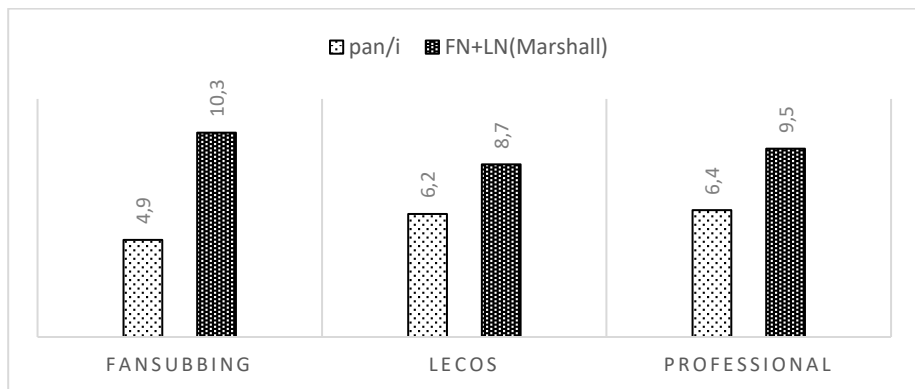


Figure 6

Incidence of *pan/I* versus first names and last name across three subcorpora in percentages

¹² The occurrences are corrected in the sense that the values for non-professional translations are divided by three to make them comparable with the single professional translation.

4. Discussion and conclusions

Given the analysis presented above, a clear picture emerges. The amateurs promote American addresses, regardless of shifts in (im)politeness it entailed when translated into Polish. Following American conventions of addresses, fansubbers resorted to the first name (e.g., “Nick”) or the full name (FN+LN, e.g., “Nick Marshall”) twice as often as the formal form *pan/i* (e.g., “panie Marshall”). Fansubbers also frequently adhered to the informal *ty* address. The FN and the *T* form employed by amateurs speak for a more informal translation offered by fansubbers. The Polish system of etiquette does not allow using the *T* form with interlocutors one is not on friendly terms with, so the amateurs’ renderings are closer to the American culture and far from the Polish terms of address conveying politeness. The amateurs’ renderings thus demonstrate a tendency to foreignisation, i.e., adhering to the address system of the source language and culture (American) and not the target language and culture (Polish). This violates the politeness rules in Polish and retains the sense of otherness. Such an approach to subtitling resonates with what Nornes (1999) claims, namely, that by retaining otherness, the subtitler enables the secondary audience (i.e., one from a different country) to experience the foreign. Nornes (1999) calls for norm-defying practices in subtitling and discarding measures that lead to smoothing over the original text and making it softer for the target audience. As observed in another study (Bączkowska 2021), fansubbers who translated the subtitles under investigation seem to adhere to this claim.

On the other hand, students tapped into the formal *pan/i* form more often than the first name, which resembles the professional translation. Compared to amateurs’ renderings, they also capitalised on the *P* address instead of the full name or last name address. The form *P+LN*, as already mentioned, is not considered polite in Polish, yet it can be heard in some areas of Poland, mainly rural areas and the old Prussian lands (Miodek 1980: 178, 1991: 34, Huszcza 2006: 107). Whilst it is not a polite form and widely used in Polish, it seems to be a better choice

than using the full name address (e.g., “Nick Marshall”), which in turn is an unusual, odd and generally unacceptable way of addressing people in Polish. The FN+LN form was avoided by the students as much as the informal *ty* address, and if the *T* form was chosen, it conveyed mostly negative overtones. The students seemed to adhere to the rules of the Polish address system to a greater extent than amateurs. Moreover, the hypocoristic addresses are richer and more frequent, which is indicative of a tendency to convey politeness. The address *proszę pana/i*, the polite term of address, occurs in the LeCoS data but only marginally in the fansubbing data, which also speaks for students’ tendency to use polite and more formal terms of address.

The obvious limitation of this investigation is the number of translations involved, which spanned three amateur renderings, three students’ versions and one professional subtitled DVD version. Therefore, this study should not be generalised to evaluate the nature of subtitling *per se*. However, to the best of my knowledge, studies comparing professional, fansubbing and students’ subtitles, focusing on impoliteness, have been scarce; thus, the analysis presented here is a small-scale yet, hopefully, a valuable contribution to the general investigation of both non-professional AVT and impoliteness studies.

Overall, from the analyses presented in this paper, it can be concluded that the students’ subtitles appear to be more formal and more polite than those prepared by amateurs inasmuch as the students avail of the Polish system of politeness etiquette more often than the amateurs, who follow the rules for addresses in American English.

Acknowledgement

This study was conducted during my six-week research stay at the University of Galway, Ireland (former University of Ireland, Galway), School of Languages, Literatures and Cultures, in 2016 and 2020. I want to thank my Host, Prof. Laura Mc-

Loughlin, for her constant support and invaluable discussions on AVT.

References

- Bączkowska, Anna (2013). "Podejścia dyskursywne w teorii (nie)grzeczności". In: Maria Pająkowska-Kensik, Anna Paluszak-Bronka, Krzysztof Kołatka (eds). *Język, wielokulturowość, tożsamość*. Bydgoszcz: Bydgoskie Towarzystwo Naukowe, 19–33.
- Bączkowska, Anna (2015a). "Quantitative study of non-professional subtitles and implications for corpus-based translator training". In: Rachele Antonini, Chiara Bucaria (eds). *Non-Professional Interpreting and Translation in the Media*. Frankfurt: Peter Lang, 89–114.
- Bączkowska, Anna (2015b). "Learner corpus of subtitles and subtitler training". In: Sylvia Bruti, Elisa Perego (eds). *Subtitling Today: Shapes and Their Meanings*. Newcastle: Cambridge Scholars. 221–247.
- Bączkowska, Anna (2016). "Well as a discourse marker in learner interlingual subtitles". In: Meng Ji (ed.). *Empirical Translation Studies: Interdisciplinary Methodologies Explored*. Sheffield: Equinox, 149–179.
- Bączkowska, Anna (2019). "Apelatywne formy adresatywne w polskim nieprofesjonalnym tłumaczeniu komedii romantycznej *What Women Want*". *Rocznik Przekładoznawczy* 14: 43–72.
- Bączkowska, Anna (2020). "Healthy lifestyle, dieting, fitness and bodybuilding: compliments in the context of Polish online discussion forums and message boards". In: Maria Elena Placencia, Zohreh R. Eslami (eds). *Complimenting Behavior and (Self-)Praise across Social Media: New Contexts and New Insights*. Amsterdam: John Benjamins, 213–236.
- Bączkowska, Anna (2022a). "Forms of address in Polish non-professional subtitles". In: Barbara Lewandowska-Tomaszczyk, Marcin Trojszczak (eds). *Language and Practice in Professional Contexts*. Cham, Switzerland: Springer, 71–96.
- Bączkowska, Anna (2022b). "Explicit and implicit offensiveness in dialogical film discourse in *Bridget Jones* film". *International Review of Pragmatics* 14: 198–225.

- Bednarek, Monika (2019). *Creating Dialogue for TV: Screenwriters Talk Television*. London: Routledge.
- Biber, Doug, Stig Johansson, Geoffrey Leech, Susan, Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bousfield, Derek (2008). *Impoliteness in Interaction*. Amsterdam: John Benjamins.
- Braun, Friederike (1988). *Terms of Address: Problems of Patterns and Usage in Various Languages and Cultures*. Berlin: De Gruyter.
- Brown, Roger, Albert Gilman (1960). "The pronouns of power and solidarity". In: Thomas A. Sebeok (ed.). *Style in Language*. Cambridge: MIT Press, 253–376.
- Brown, Penelope, Stephen C. Levinson (1978). "Universals in language usage: politeness phenomena". In: Esther Goody (ed.). *Questions and Politeness*. Cambridge: Cambridge University Press, 56–289.
- Brown, Penelope, Stephen C. Levinson (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Clyne, Michael, Catrin Norrby, Jane Warren (2009). *Language and Human Relations: Styles of Address in Contemporary Language*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan (2011). *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan, Michael Haugh (2014). *Pragmatics and the English Language*. Basingstoke: Palgrave Macmillan.
- Erickson, Frederick (1982). "Money tree, lasagna bush, salt and pepper: Social construction of topical cohesion in a conversation among Italian-Americans". In: Deborah Tannen (ed.), *Analysing Discourse: Text and talk*. Georgetown: Georgetown University Press, 43–70.
- Eelen Gino (2001) *A Critique of Politeness Theories*. Manchester: St. Jerome Publishing.
- Felix-Brasdefer, Cesar (2015). *The Language of Service Encounters: A Pragmatic-Discursive Approach*. Cambridge: Cambridge University Press.
- Fraser, Bruce, William Nolen (1981). "The association of deference with linguistic form". *The International Journal of the Sociology of Language* 27: 93–110.
- Garcés-Conejos Blitvich, Pilar (2010). "A genre approach to the study of impoliteness". *International Review of Pragmatics* 2 (1): 46–94.
- Geyer, Naomi (2008). *Discourse and Politeness: Ambivalent Face in Japanese*. London: Continuum.

- Grainger, Karen (2011). ““First order” and “second order” politeness: Institutional and intercultural contexts”. In: LPRG (eds.). *Discursive Approaches to Politeness*. Berlin: Mouton de Gruyter, 167–188.
- Grice, Herbert P. (1975). “Logic and conversation”. In: Peter Cole, Jerry L. Morgan (eds). *Syntax and Semantics*. Vol. 3. *Speech Acts*. New York, NY: Academic Press, 41–58.
- Haugh, Michael (2007). “The discursive challenge to politeness theory: an interactional alternative”. *Journal of Politeness Research* 3: 295–317.
- Haugh, Michael (2015). *Im/politeness Implicatures*. Berlin: Mouton de Gruyter.
- Hill, Virginia. (2014). *Vocatives: How Syntax Meets Pragmatics*. Leiden: Brill.
- Huszczka, Romuald (2006). *Honoryfikatywność: gramatyka, pragmatyka, typologia*. Warszawa: PWN.
- Kádár, Dániel Z., Michael Haugh (2013). *Understanding Politeness*. Cambridge: Cambridge University Press.
- Kecskes, Istvan (2017). “Implicitness in the use of situation bound utterances”. In: Piotr Cap, Marta Dynel (eds). *Implicitness: From Lexis to Discourse*. Amsterdam: John Benjamins, 201–215.
- Kerbrat-Orecchioni, Catherine (2004). “Introducing polylogue”. *Journal of Pragmatics* 36: 1–24.
- Klemensiewicz, Zenon (1946). “Pan i Obywatel”. *Język Polski* XXVI/2: 33–42.
- Kretzenbacher, Heinz L., John Hajek, Robert Lagerberg, Agnese Bresin (2013). “Address forms in language contact and language conflict: the curious history and remnants of *onikani* in Czech”. *Australian Slavonic and East European Studies* 27/1: 87–103.
- Locher, Miriam (2004). *Power and Politeness in Action. Disagreements in Oral Communication*. Berlin: Mouton de Gruyter.
- Labov, William (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.
- Lakoff, Robin, Deborah Tannen (1979). “Communication strategies in conversation: the case of scenes from a marriage”. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society* 1979, 581–592.
- Lakoff, Robin (1973). “The logic of politeness; or minding your p’s and q’s”. In: Claudia Corum, T. Cedric Smith-Stark, Ann Weiser (eds). *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 292–305.

- Leech, Geoffrey (1983). *Principles of Pragmatics*. London: Longman.
- Locher, Miriam (2006). "The discursive approach to polite behavior: A response to Glick". *Language in Society* 35: 733–735.
- Locher, Miriam A., Richard J. Watts. (2005). "Politeness theory and relational work". *Journal of Politeness Research* 1(1): 9–33.
- LPRG (Linguistic Politeness Research Group). (2011). *Discursive Approaches to Politeness*. Berlin: Mouton de Gruyter.
- Lyon, John (1970). *New Horizon in Linguistics*. Harmondsworth: Penguin.
- Łaziński, Marek (2006). *O Panach i Paniach: Polskie rzeczowniki tytułowe i ich asymetria rodzajowo-płciowa*. Warszawa: PWN.
- Massidda, Serenella (2015). *Audiovisual Translation in the Digital Age: The Italian Fansubbing Phenomenon*. London: Palgrave Macmillan.
- Mills, Sara (2002). "Rethinking politeness, impoliteness and gender identity". In: Lia Litosseliti, Jane, Sunderland (eds). *Gender Identity and Discourse Analysis*. Amsterdam: John Benjamins, 69–89.
- Mills, Sara (2003). *Gender and Politeness*. Cambridge: Cambridge University Press.
- Mills, Sara (2011). "Discursive approaches to politeness and impoliteness". In: LPRG (eds). *Discursive Approaches to Politeness*. Berlin: Mouton de Gruyter, 19–57.
- Mills, Sara (2017). *English Politeness and Class*. Cambridge: Cambridge University Press.
- Miodek, Jan (1980). „Jeszcze o sposobach zwracania się do drugich”. *Język Polski* 2–3: 177–181.
- Miodek, Jan (1991). *Odpowiednie dać rzeczy słowo*. Warszawa: PIW.
- O'Hagan, Minako (2009). "Evolution of user-generated translation: Fansubs, translation hacking and crowdsourcing". *Journal of Internationalization and Localisation* 1/4: 94–121.
- Pike, Kenneth (1954). *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton.
- Pisarkowa, Krystyna (1979). „Jak się tytułujemy i zwracamy do drugich”. *Język Polski* LIX: 5–17.
- Rusiecki, Jan (1988). "Appellatives and verbal forms of address". *Linguistica Silesiana* 29: 29–42.
- Searle, John (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sikora, Kazimierz (1993). "Jak pan zawędrował na wieś". *Język Polski* LXXIII/4-5: 298–307.
- Slobin, Dan (1963). "Some aspects of the use of pronouns of addresses in Yiddish". *Word* 19/2: 193–202.

- Sosnowski, Wojciech (2015). "Formy adresatywne. Aspekt językowy i socjologiczny". In: Danuta Roszko, Joanna Satoła-Staškowiak (eds.). *Semantyka a konfrontacja językowa 5*. Warszawa: Slawistyczny Ośrodek Wydawniczy, 319–332.
- Spencer-Oatey, Helen (2008). *Culturally Speaking: Managing Rapport through Talk across Cultures*. 2nd ed. London and New York: Continuum.
- Stone, Gerald (1977). "Address in the Slavonic languages". *Slavonic and East European Review* 55/4: 491–505.
- Stone, Gerald (1981). "Pronominal address in Polish". *International Journal of Slavic Linguistics and Poetics* 2: 55–76.
- Taylor, Talbot (1992). *Mutual Misunderstandings: Scepticism and the Theorising of Language and Interpretation*. London: Routledge.
- Terkourafi, Marina (2005). "Beyond the micro-level of in politeness research". *Journal of Politeness Research* 1: 237–262.
- Terkourafi, Marina (2008). "Toward a unified theory of politeness, impoliteness, and rudeness". In: Derek Bousfield, Miriam Locher (eds.). *Impoliteness in language*. Berlin: de Gruyter Mouton, 45–74.
- Tomaszkiewicz, Teresa (2006). *Przekład audiowizualny*. Warszawa: PWN.
- Van der Bom, Isabelle, Sara Mills (2013). "A discursive approach to the analysis of politeness data". *Journal of Politeness Research* 11/2: 179–206.
- Vismans, Roe (2019). "Address negotiations in Dutch emails". In: Bettina Kluga, María Irene Moyna (eds.). *It's not all about you: New Perspectives on Address Research*. Amsterdam: John Benjamins, 253–280.
- Watts, Richard J. (2003). *Politeness*. Cambridge: Cambridge University Press.
- Watts, Richard J. (2005). "Linguistic politeness research: *Quo vadis?*". In: Richard J. Watts, Sachiko Ide, Konrad Ehlich (eds.). *Politeness in Language*. Berlin: Mouton de Gruyter, xi–xlvii.
- Watts, Richard J., Sachiko Ide, Konrad Ehlich (1992). *Politeness in Language: Studies in its History, Theory and Practice*. Berlin: Mouton de Gruyter.
- Wenger, Etienne (1988). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.
- Werkhofer, Konrad (2005). "Traditional and modern views: The social constitution and power of politeness". In: Richard Watts, Sachiko Ide, Konrad Ehlich (eds.). *Politeness in Language: Studies in its History, Theory and Practice*. Berlin: De Gruyter Mouton, 155–199.

- Zaręba, Alfred (1974). "Czasowniki określające sposób zwracania się do drugiej osoby. *Język Polski* LIV: 376–388.
- Zwicky, Arnold (1974). "'Hey, what's your name!'". *Chicago Linguistics Society* 10: 787–801.

Anna Bączkowska
ORCID ID: 0000-0002-0147-2718
Instytut Anglistyki i Amerykanistyki
University of Gdańsk
Wita Stwosza 51
80-308 Gdańsk
Poland
anna.baczkowska@ug.edu.pl

Selected aspects of sentiment analysis in the context of human communication

JOANNA REDZIMSKA

*Received 11.12.2021,
received in revised form 29.11.2022,
accepted 30.11.2022.*

Abstract

In the modern world, AI language-based solutions are more and more present in everyday life. Language users deal with chatbots or intelligent assistants (e.g., Siri or Alexa) that provide a human-machine interaction. This interaction relates not only to the knowledge of language rules but also to the way language exchange works among its users. Furthermore, such interaction definitely suggests implied meaning and cognitive aspects of the human communication. Thus, the research question of this work proposes the thesis that in the process of natural language processing, the above-mentioned aspects must be taken into account. The present article attempts to address the issue of whether the interpretation of the sentiment analysis models the cognitive background of the language interaction (human communication). Additionally, this analysis provides qualitative data that support the integrity of the NLP processes.

Keywords

axiology, sentiment analysis, implied meaning, NLP, cognitive linguistics

Wybrane aspekty analizy sentymentu w kontekście komunikacji międzyludzkiej

Abstrakt

We współczesnym świecie rozwiązania sztucznej inteligencji, które wykorzystują język naturalny, coraz częściej stosowane są w życiu codziennym. Użytkownicy języka mają do czynienia z czatbotami lub inteligentnymi asystentami (np. Siri czy Alexa), które zakładają w swoim funkcjonowaniu interakcję pomiędzy człowiekiem, a 'maszyną' (czyli rozwiązaniami z zastosowaniem zaawansowanego oprogramowania, które potencjalnie przejmuje rolę interlokutora). Wspomniana wyżej interakcja opiera się nie tylko na znajomości zasad gramatycznych i strukturalnych języka naturalnego, ale również zasad stosowania języka pomiędzy jego użytkownikami. Używanie języka, natomiast, sugeruje wykorzystanie znaczenia implikowanego oraz kognitywnych aspektów języka w procesie komunikacji. Pytanie badawcze niniejszej pracy dotyczy kwestii, jak bardzo w procesie przetwarzania języka naturalnego powyższe aspekty są reprezentowane i jaką rolę odgrywają w rozumieniu treści poszczególnych komunikatów przez odbiorcę i nadawcę. Według autorki artykułu analiza sentymentu w znacznym stopniu egzemplifikuje kognitywne podstawy językowej interakcji pomiędzy uczestnikami sytuacji komunikacyjnej, co dostarcza również danych jakościowych potwierdzających nierozłączność w/w procesów.

Słowa kluczowe

językoznawstwo kognitywne, znaczenie implikowane, analiza sentymentu, aksjologia, przetwarzanie języka naturalnego (NLP)

1. Introduction

The relationship between language and the world has long been the focus of the research of philosophers and linguists, and it is particularly exemplified by the language determinism view (Hickman 2000) that assumes that there is a straightforward relationship between language and the world that surrounds it.

This interdependence is of particular interest to cognitive linguistics, which explains, among other aspects, the function of language in naming categories by means of which language users function cognitively; these users apply cognitive categories which construe the world around them. Additionally, semantic intuition, developed in the course of acquiring and learning a language, helps users to solve problems with meaning ambiguity, finding reference or co-reference. Furthermore, the same intuition, as well as other cultural or social aspects, allow users to make judgements or express opinions that present selected axiological content.

The procedures that seem obvious to language users, as far as the use of language is concerned, pose a significant number of challenges to be met by AI solutions. To put it in other words, one of the greatest questions is connected with modelling (by means of formal expressions, numbers or algorithms) the way people think and transferring the obtained results to solutions used in natural language processing. Thus, in this article an attempt will be made to reveal if certain formal representations (or numbers) used in sentiment analysis accurately reflect a human way of thinking subjectively.

As follows, this article focuses on selected theories and aspects related to natural language (beginning with its definition and cognitive background, as well as the process of communication). Next, the discussion moves towards axiology and sentiment analysis as fields that provide and represent qualitative aspects of the language use. Thus, in the following part there is an attempt to answer the question of whether the sentiment analysis, despite being the procedure characteristic for natural language processing, is an element of implied meaning dependent on the process of cognition.

2. Language

When discussing the phenomenon of language, two key issues must be primarily taken into account, namely the definition of language that, for the majority of language users, is simplified

to the division of language into its form and meaning. The theoretical approach towards such division is introduced in the works and ideas of Ferdinand de Saussure (de Saussure 1974 [1916]: 65–70), which mark a threshold in treating this relationship between form and meaning. In his theory the core interest is in an arbitrary relationship between form and meaning (a signifier and a signified) that are part of a language convention established among its users, where the system of language is represented by *langue* and the way it is used (conventionally) is called *parole*. Although there are other possible means of expressing meaning, still among a given community language performance is based on conventionalised turns of phrase as is highlighted by Erman and Warren (2000) as well as Buerki (2020). Yet this structuralist strand of thinking suggests some amount of objectivity both in a language form as well as in the way language rules can be formulated. In consequence it also means that language representation is objective in nature and that it is possible to formulate language rules by means of objective symbols.

Such ideas are strongly advocated by Chomsky (1972) and his generative approach to language in which the focus is mostly on an objective language form that can be generated in the same way as mathematical equations. In Chomsky's (1972) opinion, meaning is of secondary importance and any ambiguity in meaning can be explained at the level of a deep structure. So, this generative approach sees language primarily as a collection of syntagms which in an effective way describe the language competence of a speaker that is based on language rules stored in the form of a mental grammar. As Chomsky (1972: 118) further points out:

The normal use of language relies in an essential way [...] on the fact that language contains devices for generating sentences of arbitrary complexity. Repetition of sentences is a rarity; innovation, in accordance with the grammar of the language, is the rule in ordinary day-by-day performance.

Thus, in effect, this approach explains how language forms are created and used by language users as well as offers a tool which allows us to look on the method by which language is represented in a standardized way.

An opposite approach to this relationship between form and meaning is represented by cognitive studies. Since the 1970s, cognitive linguistics with its followers and opponents has gained a position in the broad spectrum of theories and approaches to language. Ronald Langacker, Leonard Talmy, George Lakoff, Mark Johnson, Mark Turner, to name only a few, are frequently enumerated as scholars whose works have given grounds for this approach and they contributed enormously to its development. In their works one can read that language (a language structure) should not be studied in a purely objective way, without any insight into semantic or contextual parameters that in the opinion of the authors mentioned above exert enormous influence on the selection of language forms used/preferred by language users. Generally, language in their understanding goes far beyond language form, and following Lakoff and Johnson (2003) language is highly metaphorical because our thinking about the everyday world is also metaphor-based.

The above presented theories mark two opposite views and are provided here only to suggest how complex nature of natural language is. Yet, discussing ideas connected with natural language and artificial intelligence always raises a difficult question of how to define natural language, the answer to which offers a multitude of explanations, each making different aspects prominent. Since the major focus of this work then is on these aspects/ definitions which are crucial for application in artificial intelligence solutions, only these approaches where language can be explained by means of formal/ objective symbols (that in consequence are easily readable by algorithms) seem justified.

2.1. Communication

An attempt to define language in the context of human communication is presented Gemel (2015: 41) who suggests one of the

major functions of language (by means of which language can be defined) is to convey successfully meaning and the intentions of the sender of a message, so that this message evokes appropriate reactions on the part of the receiver. Thus, language is understood as a tool that must be used/mastered by its users who, by applying it, fulfill communicative intentions. Further, Gemel (2015) suggests that the following simplified signalling games model can be used to present this signalling convention, which at the same time ensures an intended reaction (of course remembering that each signal per se represents the amount of information rather than information itself):

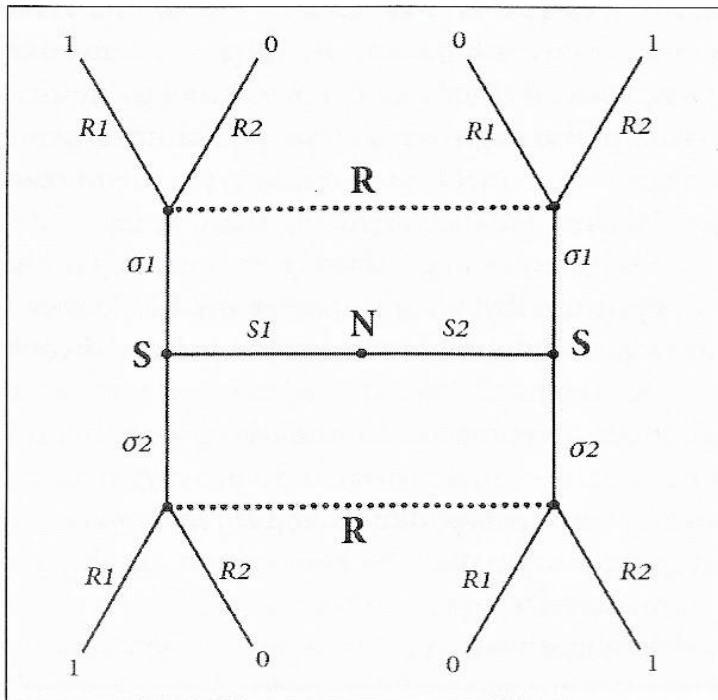


Figure 1

Adapted from Gemel (2015)

Following Gemel (2015: 41) the model presents two states (S), two signals (σ) and two reactions (R) in the form of a decision tree where each player is identified as:

N – nature
S – sender
R – receiver

Each branch of this tree exemplifies mutually excluding alternative events where:

S_1 and S_2 represent existing states of nature S_1 or S_2
 σ_1 and σ_2 represent sending a verbal signal σ_1 or σ_2
 R_1 and R_2 represent undertaking action R_1 or R_2

Thus, as Gemel (2015: 40) points out, a given (efficient) signalling convention is then based on the correlation of a reaction with an existing state on the basis of a verbal signal, which has its consequence in a positive pay-out received by both players. This pay-out can be obviously related to a given context. In the model above, one can see that only two strategies out of four give a positive reward (1) and two of them fail (0). As follows from the above explanation, language can be defined by means of a signalling convention that is explained with the application of objective symbols; the convention and symbols can be further used by algorithms in AI solutions.

Of course, the above model does not concentrate on information content that is sent by a particular signal. This issue is explained by Gemel (2015: 51) who claims that:

Signal information content can be analysed as a collection of elements constituting a mental space (points, regions formed out of points, dimensions and cross-domains dependencies). Elements positively correlated with a given signal carry information about it with a growing probability of the representation of this element in the mental space of an agent, after receiving this correlated signal. (translated by the present author).

So Gemel (2015: 52) proposes a formal representation of information content in the signalling games model (also mentioned in Skyrms 2010) that is supplemented by mental spaces and

has its representation in the form of a vector which has to take the form of a logarithm in order to make it possible to explicate situations in which a given signal does not carry any information at all:

$$\left\langle \log \frac{P(\mathbb{E}_1|\sigma_1)}{P(\mathbb{E}_1)}, \log \frac{P(\mathbb{E}_2|\sigma_2)}{P(\mathbb{E}_2)}, \log \frac{P(\mathbb{E}_3|\sigma_3)}{P(\mathbb{E}_3)}, \dots, \log \frac{P(\mathbb{E}_n|\sigma_n)}{P(\mathbb{E}_n)} \right\rangle$$

Adapted from Gemel (2015: 52)

Each informative (or communicative) act carries a given amount of information ($\mathbb{E}_{(1\dots n)}$); represents a constitutive element of the mental space (points, regions and dimensions); and can be schematically represented as:

$$\log \frac{P(\mathbb{E}_n|\sigma_n)}{P(\mathbb{E}_n)}$$

So it represents the relation of the conditional (|) probability (P) of a given state after sending a given signal that carries a given amount of information:

$$P(\mathbb{E}_n|\sigma_n)$$

with the unconditional probability (P) of this state

$$P(\mathbb{E}_n)$$

As follows from the above representation, language users perform a sequence of acts in their process of communication, and owing to their skills at processing and interpreting verbal signals, they are able to distinguish which information is relevant to them and which should be excluded.

Naturally, these skills primarily concern the ability to send and receive these signals as well as the knowledge of language.

This knowledge frequently is simplified to the knowledge of words and how to use them; yet, this knowledge of words means that users can effectively link a given language form with its meaning. Thus, language is recognized by many as a phenomenon which differentiates human beings from other beings in the world. One of the reasons for this is its complexity and especially its functions that go far beyond mere communication.

3. Axiology and sentiment analysis

It should be emphasized that, as Leibnitz claims, language is “the best mirror of the human mind” (cited in Alba-Juez and Thompson 2014: 3). Thus, analysing the intricacies of language and interpreting meaning that accompanies the language form can be helpful in comprehending human reasoning. The implied part of meaning that expresses emotional and, more particularly, subjective messages is linked to the issue of axiology in language.

According to Krzeszowski (1997: 15), there is a straightforward link between categorization and evaluation in language since they both “directly manifest themselves in language”. As Puzynina (cited in Krzeszowski 1997: 15) emphasizes, “the domain of valuative words and their meaning” and “the ways in which valuations are expressed in a language and in the structure of texts” provide grounds for linguistics and axiology to meet.

However, Krzeszowski (1997) advocates the ‘positive-negative’ distinction as the most significant for semantic analysis, at the same time refuting the ‘true-false’ distinction typical of truth-conditional logic used in the transformational-generative grammar. So, this positive-negative scale because of its generality and universality seems useful for describing and analysing semantic aspects in such a way where:

- a) following Osgood, Suci and Tannenbaum (quoted in Krzeszowski 1997), the largest part of general variance (33 %) in language is connected with the evaluation on the 'good-bad' scale
- b) the categorization based on distinguishing good from bad is learnt and implemented by children in their early years of life and it precedes their language development and language comprehension (prior to evaluative concepts such as 'ugly' or 'beautiful') since it is relevant for individual words as well as their combinations

Afterwards, it has influence on the domain of values which according to Krzeszowski (1997) is two-dimensional, where the vertical orientation reflects the hierarchy of values. The highest ones are the best ones, whereas the lowest ones represent the least desirable values. It corresponds to the model of the Great Chain of Being which explains that "the world as experienced by human beings consists of things arranged in a certain hierarchical order" (Krzeszowski 1997: 64). This model groups things into five categories that are organized in consecutive levels of the hierarchy and they are: God, humans, animals, plants and inorganic things, with God taking the highest position in this hierarchy.

This vertical organization is accompanied by the left-right horizontal one (Krzeszowski 1997) which depicts the scale of values in the form of a continuum at each level between two extreme poles: a negative one (situated to the left) and a positive one (situated to the right). Thus, as follows when analysing a concept, one looks for its category in the vertical dimension and its corresponding values in the horizontal dimension. However, the major concern of this work relates to this horizontal analysis rather than the vertical one since it represents the axiological charge of a concept.

3.1. Axiological charge

A simplified definition of the axiological charge explains it as point of reference on a positive-negative scale. According to Krzeszowski (1997), lexical items can be assigned to this scale in accordance with the following procedures:

1. eliciting absolute values of given lexical units
2. eliciting values of given lexical units as contrasted with other lexical units
3. eliciting concrete values of lexical units in given contexts

The above procedures for absolute values are accompanied by methods for eliciting the axiological charge of lexical units which, as Krzeszowski (1997) provides, are:

- a. direct rating
- b. sentence formation

As far as direct rating is concerned, it is commonly used in social sciences where respondents of questionnaires evaluate an idea by choosing one of a number of suggested options. The organization of the options follows the pattern of two polar terms with a number of options between. With larger scales (where the choice is between more than five options), it is a common practice to use various quantifiers to define precisely the level of intensity, among which the most common are 'extremely', 'very', 'slightly'.

Another method to elicit the absolute charge is directed towards sentence formation. In general, particular lexical units must be used in sentences formed by participants of the survey. These sentences are to reveal, then, their attitudes and opinions in connection with particular issues. So, participants may condemn or applaud the concept the term describes by explicitly expressing their subjective like or dislike or by placing the concept on a scale by identifying it as a positive one or a negative one. The answers to such surveys are analysed and presented

as a ratio of positive or negative sentences that make use of a given concept.

If we take into account eliciting values of certain lexical items in comparison to other lexical items, the intensity of a given property becomes evident. The surveys, as Krzeszowski (1997) describes them, are based on organising terms by participants from the most intensive ones (e.g., the most beautiful) to the least intensive ones (e.g., the least beautiful). Each position at which items are placed is given a number of points, so the final ranking presents a number of points ascribed to particular items.

When values of lexical units are elicited in specific contexts, positive and negative values are ascribed to contexts in which lexical units appear and in this way it is possible to estimate the axiological charge. The case study examples described in Krzeszowski (1997) come from SJP (Słownik języka polskiego – the Dictionary of Polish), as in his opinion monolingual dictionaries are the most valuable source of contexts. The more positive contexts are identified, the more positive a lexical unit is. However, because of its subjectivity, this method in Krzeszowski's (1997) opinion should be applied only as a supplementary strategy for finding evidence connected with a particular word and should be compared with more objective/ independent assessments.

A solution to this situation, where a monolingual dictionary is used, is offered by a balanced corpus data which provide contexts for particular words. But this method is not without flaws, since one of its disadvantages may be too much data to evaluate and annotate manually for positive or negative contexts. It is possible to have access to corpora that are annotated for sentiment but they are extremely rare and are mostly oriented towards a given category (e.g. *An Annotated Corpus for Sentiment Analysis in Political News* or *Awais Athar – Citation Sentiment Corpus*), bearing in mind that sentiment is usually marked there by assigning polarity scores to sentences or longer pieces of a text and not individual words.

3.2. Sentiment analysis

Axiological charge present in a text that is automatically evaluated is the major concern of sentiment analysis (Taboada 2016). This analysis, according to Esuli and Sebastiani (quoted in Taboada 2016) follows three main steps, namely deciding on the subjectivity of a text (if it contains subjective information that can be evaluated), then establishing the polarity of a text, and as the final step estimating what the strength of that polarity is. As Taboada (2016) points out, a lexicon-based model and a machine-learning model represent two major methods in sentiment analysis, with the lexicon-based one being the primary focus of the present work. What must be also highlighted is that from the standpoint of IT studies, this lexicon-based method seems outdated and nowadays sentiment analysis (similarly to other processes) are realized by means of machine-learning methods. Yet, in the opinion of the present author this is the lexicon-based method that in the context of human communication reveals selected cognitive aspects of language to a large extent.

3.2.1. Lexicon-based method

This approach (also called a dictionary-based method or a rule-based method) basically applies lists of evaluative words in order to identify individual lexical units in a text as positive or negative and further use this information to decide about the axiological orientation of the whole text. By the application of these lists (also called dictionaries) of evaluative words, it is possible to establish the position of a word on a plus-minus scale (i.e. if a word expresses a positive axiological charge or a negative one). Such dictionaries, as Taboada (2016) writes, may use different models whose range may start from 5,000 words (e.g. Semantic Orientation Calculator (SO-CAL) that uses a 10-point scale, from -5 to +5) to almost 76,000 words: the Macquarie Semantic Orientation Lexicon as Mohammad et al. (2009) describe. Yet, Taboada et al. (2011) highlight that the larger the dictionary, the more semantic noise it includes, thus the less

accurate the results may be. Additionally, as Taboada (2016: 13–14) remarks, one has to take into consideration in a rule-based method the role of intensifiers, i.e. “devices that change the intensity of an individual word, whether by bringing it up or down”. In order to facilitate the calculation of values affected by them, Taboada et al. (2011: 275) present the following solution:

+100 %

- really +25 %
- very +15 %
- somewhat -30 %
- arguably -20 %

This scale illustrates that calculating sentiment is not only limited to content words, but as it turns out the intensity of sentiment can be modified or even changed by various modifiers and must be taken into consideration when the total value is calculated. For example, if the opinion word ‘bad’ has a value -5 and is modified by a word ‘really’ then:

$$-5 \times 0,25 + (-5) = -6,25$$

Yet, if we change the modifier to ‘arguably’ then:

$$-5 \times (-0,2) + (-5) = -4$$

The above results show that modifiers play a significant role in presenting sentiment values and what follows in presenting one’s opinions. Another matter concerns the fact that there is a need for the most advanced lexicon-based models to include dictionaries in which there are various modifiers with assigned values listed.

Obviously, a question arises concerning the issue of negation and the ways of analysing it while processing a text. Taboada (2016:18) explains the approach in which “the effect of a negator is to shift the negated term in the scale by a certain amount, but without making it the polar opposite of the original term”.

In Asghar et al. (2017) the method of reversing the value of a negated opinion word by multiplying it by -1 is advocated, arguing that the role of negation is to tone down an evaluative phrase rather than to reverse it.

3.2.2. Applications of the sentiment analysis

Although the practical applications of sentiment analysis are many, the focus here is drawn to only a major selected aspect. As follows, it is connected with the first and probably the most obvious use that comes to mind, namely the sentiment analysis software (some call them engines) that provides a researcher with information concerning the polarity of the analysed texts of different kinds coming from different sources. Thus, this software is frequently used to analyse reviews or opinions connected with particular products and, as Taboada (2016) points out, the results of this analysis are reflected in marketing or political campaigns.

4. Discussion

A natural question that comes to mind concerns the relation between the above mentioned theories and their role for the present work. It must be highlighted again that the aim of this discussion is to reveal linguistic intricacies of sentiment analysis rather than IT procedures that are more applicable nowadays for the same sake, e.g Machine Learning. That is why, at first, the sketch of definitions of language are given in order to provide insight into the nature of language, on the one hand, and the possibility of approaches to language and its functions, on the other hand.

Moreover, these definitions are followed by an attempt to explain the nature of the process of human communication. For the sake of this discussion, only one representation is given, namely the one provided by Gemel (2015). This choice is justified by the fact that Gemel (2015) explains human communication by means of objective symbols and procedures, although

the process of human communication is to a significant extent subjective and context-dependent. Thus, as a matter of fact Gemel (2015) manages to cope with human communication performed by means of human language but defined with the application of formal representations; yet, these representations include cognitive aspects as well (mental spaces can serve as examples here), which contributes to the thesis of this work assuming that sentiment analysis or generally opinion (being highly subjective) can also have its formal representation.

As follows, the above discussion introduces the issue connected with using a natural language that obviously requires the knowledge of the language form and meanings that this form evokes. This knowledge is, then, practically applied in the process of communication, with the assumption that knowing language includes meanings that can be implied or expressed in an indirect way. Thus, AI solutions which are natural-language-oriented have to meet the challenge of dealing with both kinds of meanings mentioned previously as well as cooperating efficiently and successfully in the process of communication by means of sending, receiving and processing communication signals. Without any doubt, such solutions demand also applying/understanding cognitive bases of how a natural language functions among its users.

Ultimately, the research question centres on the issue of to what extent cognitive aspects of language use (communicative situations) can be formally represented, or in other words, whether such solutions can really imitate a human way of thinking. One of the options that replies to this question is strongly connected with sentiment analysis inasmuch as it reveals in what way axiological charge (emotions) is expressed by means of language constructions. What follows, as far as sentiment analysis is concerned, is that it makes use of cognitive aspects, since sentiment analysis actually focuses on implied meaning and communicative intentions more than only on conveying denotational meaning.

Consequently, sentiment analysis serves to process opinions which by definition are subjective in nature, and subjectivity

apart from the experiential attitude is one of the aspects of the cognitive approach (Lakoff and Johnson 2003). For AI solutions, then, it means that by applying certain mathematical formula they can effectively imitate human thinking and function cognitively well in understanding and processing meaning. Thus, in the sentence:

I find C.S..Lewis' books more than excellent.

the sentiment analysis will point to a positive charge that is implied by the combination of:

an intensifier: more than
and a content word: excellent,

which, when interpreted in a regular conversation, would mean a highly positive opinion about a given author. An additional task for the software is to respond to this comment in such a way that the flow of the dialogue is not interrupted in any way (one of the prototypes for this kind of interaction between a machine and a human being is ELIZA, in which Weizenbaum (1966) used the principle of looking for some key content words and in the flow of the conversation the computer programme responded to these words in such a way as to imitate a human-human verbal interaction). Thus, a successful human-machine interaction involves not only focusing on the language level but also on reading and analysing implied meanings.

Another challenging case for AI solutions involves situations where emotional charge is expressed by means of such strategies as, for example, irony or metaphor, where the focus is on meaning represented by a language form and cognitive categories that are activated when they are used; however, even more challenging are cases where implied meaning is coded by means of prosodic features, such as when ironic meaning depends on pitch or intentional articulation strategies that are meant to mock others. Thus, the example quoted above, if pronounced in a mocking way, may carry an implied meaning connected with

emotions (to be more precise, with attitudes) which has a completely different axiological charge (a negative one) in comparison to the one expressed in the written form. Yet, although the idea of context does not fit within the scope of this work, it must be mentioned that the interpretation of given language structures may vary depending on this context and that context itself partially represents cognitive bases of language.

5. Conclusions

As discussed above, language-based modern technologies face a number of obstacles that must be surmounted for these technologies to function effectively. Taking the research question of this work into consideration, one of these issues relates to the way in which cognitive aspects that accompany human language use are voiced in natural-language based software. In order to prove the thesis, the example of sentiment analysis is given which illustrates that AI solutions depict selected cognitive intricacies related to language use. Language modelling through sentiment analysis effectively reflects cognitive aspects of real-world modelling.

Furthermore, following the model of communication presented by Gemel (2015), it is expected that the algorithm will be able to distinguish which tracks in the communicative situation should be taken into account when processing language and which ones should be excluded. Such expectations are justified by practical reasons, namely that natural-language-oriented software, apart from facilitating/enhancing language production and language comprehension has to respond within the shortest time possible. Consequently, the human-machine interaction should be reliable as far as language form is concerned but in addition the pace of this interaction must resemble the human one as much as is achievable.

References

- Alba-Juez, Laura, Geoff Thompson (2014). *Evaluation in Context*. Philadelphia: John Benjamins Publishing Company.
- Asghar, Muhammed Zubair, Aurangzeb Khan, Shakeel Ahmad, Maria Qasim, Imran Ali Khan (2017). "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme". *PLoS ONE* 12/2: e0171649.
- Buerki, Andreas (2020). *Formulaic language and linguistic change*. Cambridge: Cambridge University Press.
- Chomsky, Noam (1972). *Language and mind* [enlarged edition]. New York: Harcourt
- de Saussure, Ferdinand (1974) [1916]. *Course in general linguistics*. London: Peter Owen.
- Erman, Britt, Beatrice Warren (2000). "The idiom principle and the open choice principle". *Text* 20/1: 29–62.
- Gemel, Aleksander (2015). *Językowy model poznania. Kognitywne komponenty w kontynentalnej filozofii języka*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Hickmann, Maya (2000). "Linguistic relativity and linguistic determinism: some new directions". *Linguistics* 38/2: 409-426.
- Krzeszowski, Tomasz (1997). *Angels and Devils in Hell: Elements of Axiology in Semantics*. Warszawa: Wydawnictwo Energeia Race Jovanovich.
- Lakoff, George, Mark Johnson (2003). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Skyrms, Brian (2010). *Signals, Evolution, Learning and Information*. Oxford: Oxford University Press.
- Taboada, Maite (2016). "Sentiment Analysis: An Overview from Linguistics". *Annual Review of Linguistics* 2. Retrieved January 4, 2020 from https://www.researchgate.net/publication/283954600_Sentiment_Analysis_An_Overview_from_Linguistics
- Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberley Voll, Manfred Stede (2011). "Lexicon-based methods for sentiment analysis". *Computational Linguistics* 37: 267-307.
- Weizenbaum, Joseph (1966). "ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine". *Communications of the ACM* 9/1: 36–45.

Joanna Redzimska
ORCID iD: 0000-0001-7837-9397
Uniwersytet Gdański
Instytut Anglistyki i Amerykanistyki
ul. Wita Stwosza 51
80-308 Gdańsk
Poland
joanna.redzimska@ug.edu.pl

How concepts are understood – online collaborative work and corpus data

BARBARA LEWANDOWSKA-TOMASZCZYK

*Received 1.03.2022,
received in revised form 19.12.2022,
accepted 20.12.2022.*

Abstract

The present study examines the engagement of Polish university students of English and American university non-language majors in a virtual exchange within the TAPP project within a semester of their studies in the year 2019 aimed at achieving the clarification of the senses of colloquial language as well as terminological yearnings the American students use in their written works. The paper discusses the concepts of *broad meaning* and *linguistic naturalness* and introduces a cognitive linguistic interpretation of interlanguage *commensurability* in translation. The materials researched in the present paper include samples of the students' online queries, peer correction, supplemented by a presentation of those lexical forms which needed further lexicographic and corpus data consultation.

Keywords

broad meaning, corpus data, commensurability, dictionary, native-language, naturalness, non-native language, online peer collaboration, TAPP

Jak rozumiemy pojęcia – współpraca online i dane korpusowe

Abstrakt

Studium przedstawia zaangażowanie polskich studentów anglistyki i amerykańskich studentów kierunków nie-językowych we współpracę online w ramach projektu TAPP w ciągu jednego semestru studiów w roku 2019. Bardziej szczegółowo, praca prezentuje sposoby zastosowane przez polskich studentów dla uzyskania głębszego wyjaśnienia zarówno znaczeń potocznych jak również terminologii używanej w esejach amerykańskich studentów. Artykuł objaśnia pojęcia *znaczenia szerokiego* i *naturalności językowej* oraz wprowadza znaczenie *współmierności międzyjęzykowej*, znane z prac językoznawstwa kognitywnego. Materiały prezentowane w pracy zawierają próbki wzajemnych tekstów i poprawek oraz pytań i odpowiedzi, jakie studenci wymieniają między sobą, uzupełnione poprzez omówienie tych form leksykalnych, które wymagają dalszego pogłębiania treści i formy w materiale słownikowym oraz w korpusach językowych.

Słowa kluczowe

dane korpusowe, język nierodzimym, język rodzimy, naturalność, Projekt Transatlactic Pacific (TAPP), słownik, współmierność międzyjęzykowa, współpraca online między rówieśnikami, znaczenie szerokie

1. Introduction

The successful use of native – non-native team-based work depends on a number of factors such as, primarily, sufficient proficiency in the common language, in our case English, the ability of team members to work together collaborating on the assigned cases, relevant intercultural as well as some digital competences both to smoothly exchange information with the partners as well as to widen the repertory of knowledge needed to fully comprehend the language used by the partners. The patterns of

collaborative work during the project life time (one semester), will be presented together with the types of corrective feedback. Further exploratory materials, aimed to lead to the identification of fuller and more precise meanings both of the terminological uses in the native speakers' works as well as the colloquial language the US peers used in their less formal style. This study identifies comprehension problems the Polish students experience during the exchanges. What follows is a presentation of materials to support solving their comprehension problems, in particular online dictionaries, language corpora – such as the Corpus of Contemporary American English, contrasted with the British National Corpus (BNC) and the most up-to-date NOW corpus available online. Additional tools are the National Corpus of Polish (nkjp.pl) and the parallel English-Polish corpora publicly available to support Polish students' translation tasks. Our data suggest that future studies could complement the analysis with such data derived from other channels (e.g., informal online blogs or posts), and can be used by students for communication and organization of native – non-native student collaborative teamwork.

2. Collaborative activities and the technological turn

It is widely accepted (e.g., Surowiecki 2005, Lewandowska-Tomaszczyk and Bogucki 2016) that groups working as teams are more effective than efforts of an individual. In the context of Second Language Acquisition, this principle seems to work particularly well and collaborative tasks are especially strongly inspired by a technological turn in computer science and development of computer language tools. Moreover, both emailing, video-meetings and communication platforms of all types are widely used at present as educational tools. Learning is mediated then by such a context in a project discussed in the present study. The students not only exchange the texts and introduce their comments but also ask questions to understand interpretative nuances of the texts, which contributes both to their subsequent translation tasks but also to the raising both of their

linguistic and inter-cultural awareness. In this sense, the study analyzes the materials which combine the processes of conscious reflection with those of writing skill development with the American students and moreover, equally significant, translation skills with the Polish students.

3. TAPP: US-Polish university students cooperation project

The cooperation referred to in the present paper is a project within a larger Trans-Atlantic & Pacific collaborative scheme (TAPP). It was first started in the 1999–2000 academic year by Bruce Maylath and Sonia Vandepitte and it involves the collaboration of students in two or more universities in different countries (Humbley et al. 2005). The TAPP has grown to include 41 universities in 19 countries (see TAPP site <https://sites.psu.edu/massimoverzella74/what-is-tapp/>) and its main aim is “to share insights into collaborative writing across borders and cultures, and, in the course of this work, to gain knowledge of others’ cultural bases.”

The engagement of Polish university students of English and American university non-language majors in this virtual exchange within the TAPP project started in the year 2015. The research published around this topic ranges from a discussion of the students’ correction strategies, dynamic profiles, reflective memos, as well as the use of argumentative structures in their exchange (Lewandowska-Tomaszczyk 2021a, Lewandowska-Tomaszczyk and Słomski 2016, Lewandowska-Tomaszczyk and Veramoothoo 2021). The present study analyses ways the Polish students collaborating with American students within the TAPP project achieve the clarification of the senses of lexical and terminology items the American students use in their written works, consulting the vague and unclear expressions in lexicographic materials, online documents, and authentic corpus data.

The collaboration varies among pairs or groups of instructors, but typically consists of US-based students in technical and professional writing classes creating documents, and students abroad translating these texts into their national languages, providing feedback on each other's work, or engage in both commenting, and eventually translating them into Polish in the case of Polish students.

In the cooperation stage referred to in this paper, the universities cooperating were the University of Minnesota and University of Applied Sciences in Konin (Poland). The Minnesota students were all non-language students, while the Konin students who took part in the collaboration were two groups of first and second year students of the MA programme in Translation Studies. Their curriculum combined theory and language practice in general English and in the respective domains, so the TAPP cooperation provided an opportunity for them to use and develop their language skills. Essentially, the TAPP collaboration between the Polish and US students functioned as a peer review activity in which the groups received feedback, usually shared by email messages, on their essays from their TAPP partners and they often used Google docs, PDFs and Word documents to provide their feedback. As mentioned above, the Konin students were additionally required to provide English-to-Polish translation of their US partners' texts.

The lecturers introduced this project by discussing terminology and concepts referring to various genres and styles. The students completed the first draft of their essays (in some cases – their research-based proposals, in some others – more personal notes concerning their first years at the university) and then emailed them to their partner students. They commented on their peers' texts, asking for clarification on terms and offering grammatical and syntactical suggestions. The students answered with their suggestions and either made the suggested changes or explained in an email or in a videocall why they had written the particular passage. If both parties were satisfied, the exchange ended there. However, many pairs continued the process with further drafts of the text until full satisfaction was

achieved. The Polish students were then asked to present the translation of their partner students' text to the Polish instructor.

4. Reflective learning

At each of the stages of the Poland-US cooperation Polish students submitted their reflective learning memos (Lewandowska-Tomaszczyk 2021), a practice carrying relatively high educational value, which has been used in such contexts for some time. Its popularity has grown to the extent that it is often adopted by educational practitioners who teach classes of this type (e.g., Calderhead 1987; Allard et al. 2007).

The data, collected here, indicate that such a cooperative format as the one applied here as well as corpus-based data enriched with supplementary online materials, involve students performing tasks which can be believed to promote reflection. The materials available here are written data coming originally in the form of student exchange of email letters and notes and online discussions with their peers and with the teachers coordinating this project. Reflective discussions and learning refer in this context to contemplating the content in various contexts (Farr and Riodan 2012) as well as interpreting and discussing them in classroom contexts.

5. Native – non-native peer interaction

The success of cooperative education programs depends upon the mutual benefit of the students and lecturers. Qualitative case studies of this kind provided some evidence that this a cooperation format produces fairly high levels of student performance and skill development.

TAPP collaborations have commonly included writing-translation projects, in which one class writes texts in a source language, then works with another class that translates these texts into a target language (Maylath et al., 2013, 2013a). The exchange referred here is based on writing tasks performed by US

students (NS) and the translation of these texts from English into Polish with the translations commented upon by the Polish students from the same group, and modified accordingly, when needed. Apart from general structural and syntactic changes, the students made an attempt to provide clarification of some of the less common terms as well as idiomatic and colloquial expressions the American students used in their texts. This part of task is invariably related with conceptual – and linguistic – indeterminacy of linguistic meanings.

6. Concept ambiguity and clarification strategies

It is generally assumed in Cognitive Linguistics, which is taken as the basic linguistic framework in the present study, that meanings, which are defined as convention-based conceptualizations, are not discrete entities, fully determined, even in fuller context.¹ Therefore, as argued in Lewandowska-Tomaszczyk (2021) it is essential to identify first the basic, *prototypical senses* of concepts and then their *broad meanings*, which include, apart from the core part, their contextual, culture-specific, and connotational properties defined in terms of a parametrized set of their system-related *semasiological* as well as *onomasiological* properties (Geeraerts 2015), emphasizing the significant role of extralinguistic reality in the process of naming. Thus it is also needed to adjust the study methodology towards a multifaceted analysis of linguistic forms and consider the interdisciplinary – linguistic, cultural and social domains – to identify the *cultural conceptualizations* of the analysed forms (Sharifian 2017). In the present case a cognitive corpus-based analysis in monolingual English contexts and in the translation data of lexicographic and parallel corpus materials used to be performed in order to more fully determine the relevant linguistic meanings.

¹ This is particularly true of context-free abstracted, less specific, lexical meanings in which *schematization* plays a role, when contrasted with what Langacker calls “*usage events*, i.e. the actual pronunciations and contextual understandings” (Langacker 2008: 16).

In the context of the cooperation discussed in the present paper, a complex set of areas of reference in cognitive cross-linguistic analyses of what is considered *broad linguistic meanings* (Lewandowska-Tomaszczyk 1989, 2012, 2012a) is carried out in terms of the following *language forms clarification strategies*:

- (i) the exchange of the students' authored texts between the American and Polish students
- (ii) identification of problem areas, typically connected with the semantic scope and uses of particular problem areas in the partners' texts
- (iii) searching up monolingual national and general English corpora as well as the materials of the English-to-Polish and Polish-to-English translational corpora (*paralela* Pezik 2014) for particular lexis and structures, their frequencies and collocations (Pezik 2012)
- (iv) identification of relevant dictionary entries in online dictionaries to examine the scope of use, genre and style of particular language chunks
- (v) identification of Polish equivalents in the parallel corpus data
- (vi) searching up bilingual dictionaries to validate the Polish equivalents
- (vii) identification of and reflection over the Polish uses of the equivalents to consider them as candidate equivalents in comparable texts identified in national corpora and over the web-based publications
- (viii) validation of the equivalents and the Polish texts in the translation produced by the translator trainees (student) for their adequacy and naturalness in a specific genre and style (reflecting on the broad senses of particular forms in context)
- (ix) an optional written task report on the performed procedure and the achieved final result in terms of the particular meaning elucidation
- (x) use of the Polish forms in various relevant contexts.

6.1. Naturalness

Apart from the understanding of the concept of *broad* meaning, the notion of *naturalness* also deserves foregrounding, particularly in the translational setting.

The broad senses are the senses that involve, apart from their core – sometimes criterial, parameters, the position of the particular meaning within a system of interrelated concepts, i.e., its semantic relational meanings and forms, activating polysemy, synonymy, and antonymy links, together with their hyponymically and hyperonymically related forms, the word usage syntactic criteria in natural language contexts, and, last but not least, their cultural-discourse adequacy. In other words, both lexical semantic criteria, as well as the syntactic adequacy and discourse type- and culture-bound appropriateness play a role. Corpus materials enable the learner to scrutinize these qualitative semantic aspects in a broad range of texts.

Corpus data relevant for translation tasks are both of the qualitative and quantitative kind. Frequencies of use and quantitative distributional facts (e.g., collocations) related to contextual factors can provide new insights into contrastive language studies and open up new prospects in translation theory and practice. The frequency characteristics will also illuminate a qualitative factor with respect to the examined data, namely, the degree of *naturalness* associated with individual constructions.² John Sinclair, who first discussed the concept of naturalness on language in (1983), gives the example Eng. *Prince Charles is now a husband* as an instance of low idiomaticity, low neutrality and high isolation, cumulated in the low naturalness of the sentence. Thus the concept of *naturalness* is understood as a system of the speaker's/writer's preferences of the use of a language unit, which is expressed via the frequency of its occurrence not in general language but in a specific, well-defined type of context (Lewandowska-Tomaszczyk 2001: 178). A more natu-

² For a thorough discussion of naturalness and the interpretation of *meaning in translation* consult the chapter by Lewandowska-Tomaszczyk *Meaning in the Encyclopaedia of Translation* (2022), publicly available online.

ral unit/structure then will be the one used more frequently in a given context.

The research task involving a *cross-linguistic comparison* is thus built around identifying cross-language similarity as a *dynamic notion*, represented as a cline exhibiting a gradual increase in diversification. The *degree of equivalence* between SL and TL structures can thus be measured in terms of the reference categories mentioned above, such as the typology of the category of *naturalness*, as well as categorization levels, prototypicality, image-schemata and their extensions, profiling and construal relations of various types, familiar from the Cognitive Linguistics paradigm (Lakoff 1987, Langacker 1979/1881).

The concept of *cross-linguistic cluster equivalence patterns* (Lewandowska-Tomaszczyk 2017), appeared particularly useful in collaborative corpus-informed teaching tasks at MA translation classes.

In section 8 of the present study parallel examples of contrastive cognitive-structural parameters of discourse and a corpus-based cognitive analysis of selected forms in English and Polish meanings are presented with a few examples of relevant cluster equivalents in Polish (Lewandowska-Tomaszczyk 2017). It is claimed that to uncover areas of analogy and difference cross-linguistically it is considered necessary to identify and analyse both a parametrized set of their semasiological as well as onomasiological properties (Geeraerts 2015), i.e., both the inherent meaning as well as the naming processes of a particular part as perceived in the outside world. To contextualize the study, the data obtained from such corpus materials should be discussed with students with reference to their cultural context (e.g., Nora 1992, Hofstede 1980, Trompenaars and Hampden-Turner 1997) as well as to particular styles and genres. These aspects are particularly well developed in linguistics and translations studies (e.g., Snell-Hornby 2006, Tirado 2019, Sharifian 2017).

6.2. Commensurability criteria

The criterion of *use*, listed by George Lakoff (1987) as one of these called *Commensurability Criteria* for comparing different languages is immediately relevant to translation tasks. It refers to a distributional range of particular language elements, i.e., the extent to which the range of use of such lexical elements in English as e.g., *to sit* in *Mary is sitting in this armchair* but also *The cup is sitting on the shelf*, correspond to different verb uses in other languages. In the example *[The products] may sit together on the shelf, and the consumer may think that those marked with CE are better than the others*³ the forms *sit* in these examples will correspond to *stand* or *lie* in other languages (e.g., Pol. produkty *stoją* ‘stand’ or *leżą* ‘lie’ na półce ‘on the shelf’).

The *framing* criterion combines the linguistic knowledge with the knowledge of the outside world. Different *object* or *event frames* or *schemata*, which regulate a top-down perspective on individual meanings are used in different languages e.g., in English the preference on the menu list is to treat some vegetables as individual entities used in the plural form e.g., the use of *carrots and peas* in the plural form in the English phrase *casserole with ground beef, **carrots**, and **peas***, while users of other languages (e.g., Polish) perceive them as a mass and use the singular (generic) noun in such cases (Pol. *z marchewką i groszkiem* lit. ‘with carrot and pea’).

Finally, the *organizational* criterion reflects distinct cross-linguistic perspectives on objects within a given category as in the cases of polysemy. Such cases represent distinct conceptual organization within semantic-conceptual categories across languages (Lewandowska-Tomaszczyk 2007). This criterion is evident in the case of *conceptual* or *lexical gaps*. In some languages as in English, for example, a fully lexicalized concept of *hubris* – negative pride – is absent, while it is present in other languages (e.g., Pol. *pycha* ‘hubris’ versus *duma* ‘pride’). Such and other

³ eur-lex.europa.eu

cases of commensurability *deficits* or *asymmetries* cause *meaning re-conceptualization* (Lewandowska-Tomaszczyk 2010) across languages and are both evident and discussed by the students in the analysis of translated texts.

It goes without saying that the students of translation specialism, who are the Polish subjects in the TAPP exchange, are typically familiar with these phenomena and recognize the types of *re-conceptualization processes*, connected with inherent *meaning approximation in communication* (cf. Lewandowska-Tomaszczyk 2010) which involve both changes in the *content of linguistic units* from one language to another but are also embedded in the *constructional properties* of language, i.e., its syntax and morphology (Goldberg 1994). The differences can also reside in the perception components and influence a possible *construal* of a scene (Langacker 1991: 551).

7. Peer corrective feedback

Corrective feedback the students are first engage in is understood here as “a more competent speaker’s reaction to learners’ ill-formed output” (Panova and Lyster 2002), and is considered an important facilitative strategy in second language acquisition and language development. The native – non-native student mutual feedback we have been observing in the analyzed context, has so far been a less frequent research topic in this area. Feedback can be understood either as *corrective recast* (direct corrections and providing correct forms) or *metalinguistic feedback* (applying a more descriptive strategy) (Monteiro 2014). Each strategy has its own advantages and drawbacks. It may be interesting to note that direct corrections were avoided by either group in our study, possibly due to politeness considerations, but also due to the fact that this type of interaction setting of basically the email exchange type provides a context which is more conducive to the metalinguistic rather than corrective recast strategy of feedback, making the former a more convenient and comfortable interactional cooperation technique than the latter. In some cases there exists a possibility that the students

did not feel confident enough in their own skills/comments/feedback.

8. Peer correction in practice

The in-text corrective feedback – by both, US and Polish, groups of students – includes some direct recast corrections, although short explanatory arguments are frequently also provided. In the case of out-of-text corrections, of a more descriptive or narrative type, it is the American students that are more attentive to avoid face-threatening acts by providing, often fairly elaborate, argumentation to their (positive and negative) comments. The Polish students tend to make reference to grammar books and dictionaries (*according to Oxford dictionary* in 1.) – in other words they appeal to authority rather than to their own language competences, particularly that some of the corrections proposed by the Polish students might have been considered as arguable.

Water is one of human beings' most basic needs. A shortage of water is a problem that shouldn't be taken lightly. Studies that have been conducted in the last 15 years show that drought conditions in the Red River Valley would lead to a major water deficit. Citizens of the Valley need a plan in place for the worst case scenario. The most logical and feasible solution is to bring water from the Missouri River to the area.

The problem the Red River Valley (and most of North Dakota) faces is an overdependence on surface water as its only sustainable source. The population of the Red River Valley is steadily growing along with its economy which puts further strain on the Red River and its tributaries as a source of water. If a drought were to occur, hundreds of thousands of people

Komentarz [Rec1]: I would say "necessities". I think water is not a need, it is necessary to satisfy our different needs, it is necessary for us to live

Komentarz [Rec2]: I guess you shouldn't use contracted forms in an essay (excluding citations)– but I don't know what are the requirements

Komentarz [Rec3]: "Overdependence" – according to the Oxford dictionary is written together

Figure 1
In-text corrections by a Polish student

9. Digital materials: Corpus data and online dictionaries

Working effectively in a collaborative team is not the only outcome expected in our case but also supporting technology skill development.

In order for Polish students to widen their language competences and make their language knowledge more substantiated in use, they are required to search for broad meanings of the problematic and unfamiliar language items in online materials such as corpora of all kinds, dictionaries and other online documents.

Both national corpora and parallel ones are used. Parallel corpora of translated materials, in which two or more language components are aligned, that is, are subdivided into compositional and sequential units (of differing extent and nature) which are linked and can thus be retrieved as pairs (or triplets, etc.), are not generally frequent as teaching tools. In our case, the materials are acquired from the Polish-to-English and English-to-Polish *paralela* corpus as used for the search of cluster equivalents in the course of translator education (Lewandowska-Tomaszczyk and Pęzik 2018).

10. Collaborative corpus-informed translation teaching

To prepare for the translation tasks and achieve fuller comprehension of the English originals Polish students first discussed the English texts with their authors. However, not always was a fuller comprehension of the concepts achieved, so a task of the online materials survey and investigation was performed simultaneously to the peer comment exchange.

Apart from mutual peer correction tasks the Polish students performed analyses of the concordance materials in English and Polish monolingual corpora (British National Corpus, COCA, NOW and the National Corpus of Polish) as well as the *paralela* corpus. Furthermore, the students generated and analysed col-

locational profiles of relevant items and performed a study on the syntactic/semantic construal (Langacker 1987) and reconceptualization types (qualitative analysis) in selected classes of example (Lewandowska-Tomaszczyk 2010). There are also corpus-based *grammars of English* published (Lewandowska-Tomaszczyk 2004) and widely used (e.g., Biber et al. 1999), recommended to students as reference materials. This was particularly helpful in meaning clarification and fuller comprehension both of the direct meanings as well as intended messages as well as their translational equivalents.

It may be significant to assume in this context that translational competence is not a uniform one type of competence. Rather, it can be considered as involving other types of competence (Neubert 2000) – language competence, textual competence, subject competence, cultural competence and transfer competence. It is precisely the *transfer* competence, including pattern-matching competence between a SL and a TL, accompanied by a decision-taking competence as well as performance competence, i.e., ability to perform in consequence of pattern-matching ability and conscious decision making, that are the most significant criterial properties in translation skills.

11. Lexis and terminology meaning clarification

The present section presents authentic instances of the Polish students' inquiry about particular monolexical and phraseological forms in American English the meaning of which was either unfamiliar or not clear to the Polish students, even in their contexts of use and often following explanatory or definitional attempts performed by the US students.

Apart from grammatical – syntactic and morphological operations observed in usage-based contexts (as e.g., in the case of part-of-speech *conversion* as in *She did really well at her first **meet***) – the American students used a number of lexical items, either less familiar to or less frequently used by the Polish students. They are exemplified in the contexts below. The Polish students typically ask questions of the form:

Pol. S [Q: (could you tell me/could you explain to me) what are X or Y ?]

The American students found such questions problematic to provide more exhaustive definitions. Instead, they typically gave examples, and the Polish students performed an explanatory task and a reflective part concerning the results of their search, following the answers they received.

Some of the cases semantically opaque for some of the Polish students were idiomatic expressions of various kinds as exemplified in the section below.

11.1. Idiomatic expressions (italicized in the examples)

Conventional Metaphor

(2) Sometimes the *best strategy for getting on track* is a little bit of parental guidance. My mom had my Blackboard and Campus Connection passwords so she could **keep tabs on my studies**.

(3) Fargo is okay, not much to do, not much to see, you know how it is," I said trying to *tiptoe around the subject of my grades*.

Metaphor – Jocular

(4) After *this epiphany*, I had a newfound drive for being studious.

In such cases a data search either in a dictionary and/or across corpus materials was sufficient to elucidate the sense.

Some more complex comprehension cases can be identified with the use of terminology in some domain-specific texts, particularly when they were used in highly marked American contexts.

11.2. Domain-specific terms

11.2.1. Academic

Independently of the culture specificity, some such uses were additionally marked in the context of the metaphoric mapping (e.g., (5)).

(5) First semester was *wrapping up*. During *dead week* I was studying until I had all content memorized that I didn't know. I had 3 final tests, one of which was Microeconomics, my worst class. *I was on an endless grind* mastering the material I had overlooked earlier in the semester *when I was shot gunning watery Busch Light rather than even touching the material*. I took my last final and headed home to Rochester for winter break impatiently waiting on my final grades. I ended up getting nothing lower than a B on all three tests. *I accomplished a 96 % grade on the Econ test I was dreading*. A job well done! it wasn't over yet, *I still had a second semester that I still needed to jump into*.

11.2.2. Military

The military terminology use as exemplified in (examples 6–10), was frequent with a few of US students in their discussions of a current debate there referring to the free access to weapons by all adult US citizens and the activities of the NRA (National Rifle Association) in the United States, an influential gun rights advocacy group.

extreme weapons

(6) *Extreme weapons* are those that are very powerful and that serve no practical purpose, such as hunting or self-defense. *Military-grade weapons* are similar due to their ability and purpose to kill as many people as quickly and efficiently as possible.

military grade weapon

(7) One that meets the standards set by the military. It doesn't matter who makes it, as long as it meets a set of standards that they all have to meet. Now, how this is done is rather sloppy. They all are not checked. Rather, they select a batch and check it. This is done with ammunition, clothing, and other supplies. In theory, it should all be to standard. In most cases, it must be American made, though. (...)

One aspect of military grade weapons is that they have parts that are interchangeable. If one gun breaks, you can take the parts from another broken gun and fix it. Civilian weapons may or may not be able to do this. This allows military weapons to be made by any company, during war, and have them all alike. Civilians like to stick

with name brands and pay the prices. This forces them to often purchase from the maker of the weapon.

Complete text available at <https://www.quora.com/What%E2%80%99s-a-%E2%80%9Cmilitary-grade-weapon%E2%80%9D>

(8) S: What is the difference between military grade weapons and civilian?

Currently the major difference is that *military rifles are capable of firing multiple rounds per trigger pull (either burst or automatic fire) and civilian models are not*. An actual military M4 has a slightly shorter barrel than a civilian carbine. Everything else is cosmetic or ergonomic.

(9) Q: What's the difference between 'assault' weapons and sporting weapons?

Sporting is semiautomatic. Yeah, it's a semiautomatic. It's not a military grade weapon. The military uses automatic weapons. <https://www.police1.com/firearms/articles/assault-weapons-vs-sporting-weapons-whats-the-difference-CD5WOUkkmBOSy0dk/> (COCA spoken)

(10)

SERIES:

Shameless

(IMDB) (Years: 2011–: 103 episodes) Country: USA Genre: **Comedy, Drama**

Series info

A scrappy, fiercely loyal Chicago family makes no apologies.

Episode

Hope Springs Paternal (2014) (IMDB) (Open Subtitles)

S1: How much for this chair? How much you want to pay? Well, how about

S2: Cash only. Please. Dad, it's working. They're falling for it. Isn't this fun?

S1: Yeah. – You okay? – Yeah. I'm conserving my energy for Carl's thing tomorrow. Every gun's loaded, okay?

S2: So here we got your Bushmaster XM-15 semiautomatic. Premium 16-inch chrome-lined profile barrel. Sweet.

S1: What do you think about home protection? I got three little ones on the way. Safeguarding the old homestead, huh? Hey, second amendment, brother. Got to protect my freedom one bullet at a time. Okay, so...

S2: Milkor M32 MGL-140 grenade launcher. Protects a shitload of freedoms. Six-round cylinder, double-action, military grade **weapon** Missing a few parts, but... it still shoots. It's heavy.

S1: What are we talking here? How much?

S2: Well, around five grand.

S1: \$5,000?

S2: Yeah.

S1: Aw, come on, man. How much for the bat? Holy shit. I got to go.

NOW corpus

The screenshot shows the NOW Corpus interface with the search term 'military grade weapon'. The interface includes a search bar, navigation tabs (SEARCH, FREQUENCY, CONTEXT, OVERVIEW), and a list of concordance results. The results are sorted by date and source, showing various news articles from different countries and dates. The concordance text is highlighted in blue, and the search term is highlighted in red. The interface also includes a 'CLICK FOR MORE CONTEXT' button and a 'HELP' icon.

Rank	Date	Source	Concordance Text
1	22-02-16 MY	malaymail.com	Remington and the other two defendants are culpable because they knowingly marketed a military grade weapon that is "grossly unsuited" for civilian
2	22-01-17 IN	thecitizen.in	Fog app saying: "Reports suggest that this is a Psychological Ops military grade weapon and in the hands of mal-intentioned actors, it can have serious
4	21-12-28 US	Penn Live (1)	It became clear to me those bullets were fired out of a military grade weapon ," he said. "It was powerful enough to pierce reinforced glass.
5	21-05-26 NG	dailypost.ng	released from Canisters as an aerosol. # "A grenade is a military grade weapon , says multiple security experts. We invite the World to see the lies in
6	21-05-01 AU	capsulecomputers.com.au	with aspirations of upward mobility combined with an insatiable appetite for investing in military grade weapon and vehicle technology. Arms trafficking
7	21-02-16 NG	premiumtimesng.com	be peace? When did it become legal for civilians to carry a military grade weapon ? What if we all begin to carry automatic rifles in defiance of authority?
9	21-01-17 US	scpr.org (1)	the Molotov cocktails didn't ignite. The person who had a military grade weapon did not seem to fire that weapon. It was not a massacre. It was not
10	20-05-06 US	theguardian.com	Compared with all other governments. If this was an attack with a military grade weapon the UK would be fucked. Half of you would have your lips attac
11	19-08-05 US	Westport News	to shoot people could easily, so easily, get access to a military grade weapon and that much ammo," said Hannah Shows, a former high classmate who
14	19-07-29 US	News-Press NOW (2)	, # The Transportation Security Administration said in a statement that the military grade weapon was located in the man's checked luggage at Baltimor
23	19-03-15 GB	Telegraph.co.uk (8)	and the other two defendants are culpable because they knowingly marketed a military grade weapon that is "grossly unsuited" for civilian use yet hac
24	18-03-25 US	Townhall	but back to this amazing video. # She replies it's a military grade weapon , which is not an answer. Kirk aptly noted that the military also uses
25	18-02-19 GB	ITV News	int disqualify someone from owning an assault rifle, an AR-25, a military grade weapon . I don't know what does. # "Suspected school shooter Nikolas G
26	18-01-10 NG	360Nobs.com	pasture? Who armed the herdsmen with AK47 rifles? Is carrying a military grade weapon and assault rifle legal in Nigeria? Are the Fulani herdsmen actir
27	17-10-04 NZ	New Zealand Herald	how about the toxic emissions that spray from the business end of a military grade weapon ? Related articles: # The record (well, the modern record)

Figure 2

NOW corpus concordances *military grade weapon*

A discussion of the terminology use in various contexts – both in the excerpts of US Army manuals as well in a loose conversation derived from an American soap opera show, made it possible for the Polish students to better conceptualize the lexical distinctions used by their American peers.

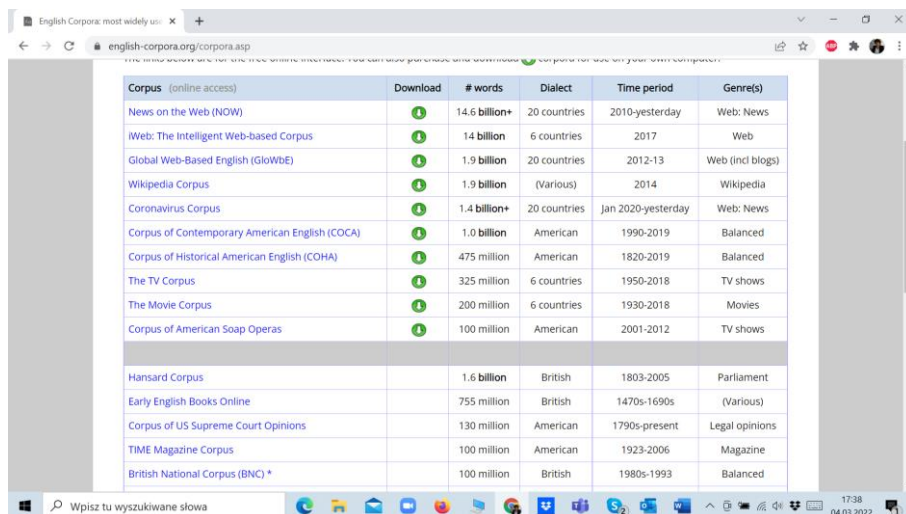
11.2.3. Sports

The culture-specificity of certain sports-related expressions was also evident in some of the US students' language.

varsity

(11) My sister had gotten into cross country; *she was running varsity* as an eighth grader.

Sentence (11) appeared problematic to the students, with reference to the expression *running varsity*. The search in a number of resources was performed by them.



Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
News on the Web (NOW)	1	14.6 billion+	20 countries	2010-yesterday	Web: News
Web: The Intelligent Web-based Corpus	1	14 billion	6 countries	2017	Web
Global Web-Based English (GloWBE)	1	1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	1	1.9 billion	(Various)	2014	Wikipedia
Coronavirus Corpus	1	1.4 billion+	20 countries	Jan 2020-yesterday	Web: News
Corpus of Contemporary American English (COCA)	1	1.0 billion	American	1990-2019	Balanced
Corpus of Historical American English (COHA)	1	475 million	American	1820-2019	Balanced
The TV Corpus	1	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	1	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	1	100 million	American	2001-2012	TV shows
Hansard Corpus		1.6 billion	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced

Figure 3
Corpora of English (selection)

noun: **varsity**; plural noun: **varsities**

1. dated, British

university.

"he had his hair cut when he got back from varsity"

2. British

(especially of a sporting event or team) relating to a university, especially Oxford or Cambridge.

modifier noun: **varsity**

"a varsity match"

3. North American

the principal team representing a high school or college in a sport or other competition.

"Miller promoted him to the varsity for his sophomore season"

The screenshot shows the 'Corpus of American Soap Operas' interface. The search results are displayed in a table with columns for rank, year, source, and context. The word 'varsity' is highlighted in the context of each result.

Rank	Year	Source	Context
1	2011	AMC	feel like I'm in high school again? Griffin: I would lend you my varsity jacket if I had one. Kendall: Actually, I'll settle for just wearing
2	2011	GH	'm supposed to be. Like if I get into Harvard and row in the varsity crew, that somehow Pentonville and all that crap that happened to me will somehow
3	2010	ATWT	I That's impressive. Tom: Hey. Also, on the baseball team, varsity shortstop. He's -- he's thrilled. So am I. Casey:
4	2010	OLTL	David: Kim, you are what we in the trade call a jgd -- junior varsity gold digger. It's a technical term, not a compliment. You're
5	2010	OLTL	my next diamond-encrusted toe ring when you should be thinking how will a vgd -- varsity gold digger -- like David Vickers Buchanan take that toe ring away from me.
6	2010	OLTL	did you go to prom with? Layla: Taylor Richards asked me. He played varsity basketball, and half the girls in school were in love with him. But
7	2010	OLTL	about basketball, though? Now, you were planning on going out for the varsity. That's a huge responsibility, you know. Matthew: So I'll work
8	2010	OLTL	We're going steady. Does this mean I need to pull out your old varsity jacket? Rex: Oh, is that what all the hot college co-eds are wearing
9	2009	ATWT	?! Mark: Because I have baseball tryouts. It's my chance to make varsity. Liberty: Oh, do you think that we can go the next day?
10	2009	ATWT	Yeah, I did, jack. Congratulations! Jonesy, my son just made the varsity baseball team. All right. Come on, let's go celebrate. I
11	2009	BB	why I've been shoved in a corner? Ridge and Rick, they're varsity. I'm J.V. It's okay for emergencies... Eric (Stammering) Thorne: But
12	2009	OLTL	You were so angry. Todd: Because of you, I didn't get my varsity letter. You messed me up royally. Mary: I couldn't breathe. My
13	2008	ATWT	innocence, okay? You concentrate on being you, huh? Like making the varsity baseball team. Carly: Yeah, really, Sweetie. Just try to get your
14	2008	ATWT	's not gon na work, unless you're looking to score with the entire varsity hockey team. Janet: Oh, zip it. You're no fun. Liberty:
15	2008	GH	team as a cover so that she could see Dillon. She ended up making varsity. The debate team -- now, that was George's thing. I got
16	2007	OLTL	I mean, you can't be surprised. This is a small town, varsity football is huge. People were pissed when you got kicked off the team --

Figure 4
Corpus of American Soap Operas *varsity*

The screenshot shows the 'Corpus of Contemporary American English' (COCA) interface. The search results are displayed in a table with columns for rank, year, source, and context. The word 'varsity' is highlighted in the context of each result.

Rank	Year	Source	Context
1	2012	BLOG shefinds.com	3267861 Geometric Jewelry You Don't Need To Be A Varsity Mathlete To Rock # You may think you left high school math in the past
2	2012	BLOG blog.northjersey.com	Pretty sure her AA was 35.875 and the corrections have been made. Search Varsity Aces # Search for: Subscribe to Varsity Aces # Enter your email add
3	2012	BLOG blog.northjersey.com	have been made. Search Varsity Aces # Search for: Subscribe to Varsity Aces # Enter your email address to receive notifications of new posts by email.
4	2012	BLOG kateharding.net	so it's not like I'm hiding my body! # I miss my varsity karate club. Our senses didn't care about what shape or size we were
5	2012	BLOG aintcool.com	- while DDL is a magnificent artist - that method acting shit is kinda junior varsity. if he can't get into character at the drop of a hat at
7	2012	BLOG xojane.com (1)	the cover of a Nordstrom's Training Bra catalog. Another was on the Varsity Surfing Team. At that point my hair had begun to grow out in a
8	2012	BLOG xojane.com	punch line. # But I made it to top three! Bri and the varsity surfer flanked me on both sides. # For top three, they put you
9	2012	BLOG youthhouseoflove.com	want a letterman's (letterperson's?) jacket, so we put her varsity letter in a shadow box. I used her school colors as a background.
10	2012	BLOG chelmsford.patch.com	Regatta this weekend, placing 19th out of 85 boats in the men's youth Varsity 8 race. Daniel Levine, Austin Drake and Johnny Coddairre all row out of
11	2012	BLOG sanmarinonotribune.com	Titan tennis supporters came true last week as the San Marino High School girls' varsity tennis team overcame a semifinal scare from visiting Claremont
12	2012	BLOG burbankbeyond.com	Pendleton added 10 yards on eight rushes. # Williams finished his first year on varsity just 63 yards shy from breaking the single season rushing record
13	2012	BLOG ...mondconfidential.org	is good, our JV team is good, but somehow when they get to varsity it just goes all downhill. " # Green said he thinks that's because
14	2012	BLOG radaronline.com	... was repeatedly bullied after she'd had sex with four of the school's varsity football players at a party at one of the player's homes Saturday night,
15	2012	BLOG suathletics.com	3.1 mile time. # In the women's open four, Syracuse's first varsity boat overcame heavy winds to emerge victorious. Led by seniors Allison Todd (coxswa
16	2012	BLOG suathletics.com	# Finishline right behind them. In the second place, was Syracuse's second varsity four. I like their teammates, the second varsity four crew turned in a q

Figure 5
COCA *varsity*

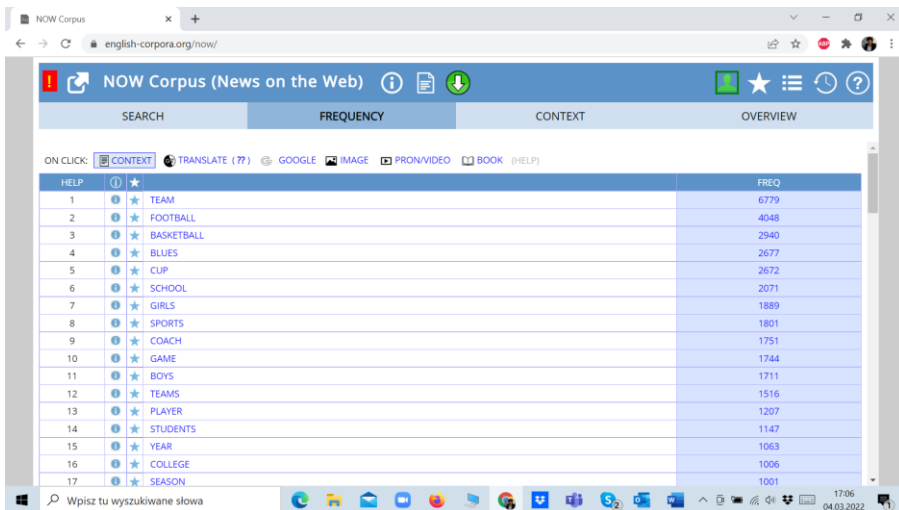


Figure 6
NOW *varsity*

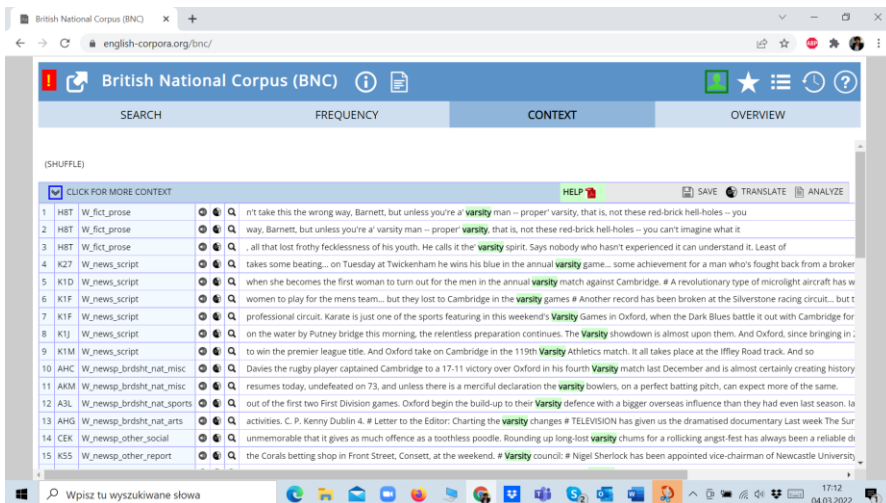


Figure 7a
BNC *varsity* collocates (i)

The screenshot shows the British National Corpus (BNC) website interface. The 'FREQUENCY' tab is selected, displaying a table of collocates for the word 'varsity'. The table has three columns: 'HELP', 'COLLOCATE', and 'FREQ'. The data is as follows:

HELP	COLLOCATE	FREQ
1	MATCH	16
2	CENTURIES	2
3	GAMES	2
4	GAME	2
5	RUGBY	2
6	WEEKEND	2
7	MAN	2
8	MATCHES	2
9	YEAR	1
10	LIFE	1
11	GROUND	1
12	GARDEN	1
13	VIRTUE	1
14	VICTORY	1
15	UNIVERSITY	1
16	UNDERGRADUATE	1
17	UNDERDOGS	1

Figure 7b

BNC *varsity* collocates (ii)

An interesting conclusion from the discussion of the polysemy of the word *varsity* was drawn by the students with reference to Figure 7., in which the British English use of the word shows both semantic types of the collocates, one referring to the university, most often to Oxford and/or Cambridge, the other – similar to the American sense – referring to sports contexts.

The area related to finances and economy was one of the more lexically problematic areas, together with the field of law, the Polish students had some comprehension problems with.

11.2.4. Finances

bidder

(12) To a veteran like me 'military-grade' means **built by the lowest bidder, with the cheapest materials**, and probably ill-fitting. If 'military grade' was so great I wouldn't have bought so much non-issue kit during my time as a soldier.9 May 2020

The identification of different *types of bidder* and their possible cluster equivalents patterns in Polish was performed in the *paralela* corpus.

(13)

In particular, bids are frequent and often small in volume (only [...] [1] of [...] [1] tenders submitted by Siemens are larger than EUR [...] [1] in size), and products are highly customised and significantly differentiated. In addition, for larger contracts, there is ex-ante uncertainty about the actual value (i.e. profitability) of a project for the winning bidder. The expected value of the price offered by the lowest bidder is therefore bound to increase as the number of **credible bidders** decreases. Hence, Siemens/VA Tech's high combined market share, the relatively small size of the remaining competitors and the elimination of an important bidder increase the possibility that a dominant position will be created as a result of the merger. [Acquis comminatoire]

Przy tym należy zauważyć, że przetargi odbywają się często, ich przedmiotem są produkty znacznie zróżnicowane i dostosowane do indywidualnych potrzeb klientów, a wartość przetargów jest niska (tylko [...] [1] z [...] [1] ofert przedsiębiorstwa Siemens miała wartość wyższą niż [...] [1] EUR). W przypadku zleceń o znacznej wartości oferent, któremu udzielono zlecenia, początkowo nie jest pewien konkretnej wartości (tzn. rentowności) przedsięwzięcia. Spadek liczby wiarygodnych oferentów powoduje jednak wzrost najniższej oferowanej ceny. W związku z wysokim łącznym udziałem w rynku przedsiębiorstw Siemens i VA Tech, stosunkowo nieznaczną wielkością pozostałych konkurentów i wyłączeniem istotnego oferenta, wzrasta prawdopodobieństwo, że koncentracja przyczynia się do stworzenia pozycji dominującej. [Porządek prawny UE]

<http://paralela.clarin-pl.eu/#search/pl/-1/lowest%20bidder/-1/0/20/0/true/0/true/-1/-1/-1/-1/source>

The corpus included examples with a number of modified uses of the noun *bidder*: *lowest bidder*, *winning bidder*, *credible bidder*, *important bidder*.

The range of Polish equivalents of the former use of the *lowest bidder* acquired from the same source includes:

(14) podmiot oferujący najniższą cenę
najtaniej z oferentów

najtańsza firma
najtańszym kosztem

which correspond to the English ‘the subject offering the lowest price’, ‘the cheapest of the offering parties’, ‘the cheapest firm’ or – an implicational rather than prototypical equivalent – ‘with the lowest cost’.

The thorough search and reflection over both meanings and uses of the lexical item made it easier for the students both to engage in the discussion with their US partners as well as to be more effective in their translation tasks.

11.3. Colloquial language

The use of colloquial, let alone vulgar, language by American students is obviously not common. However, the topics focusing on personal experiences, life stories, etc., seemed to encourage the use of such expressions, which, in more academic types of text, would be considered unacceptable.

The colloquial language poses some problems for the Polish students. Therefore, the reflection and close scrutiny of such expressions is part of the classroom tasks, focusing on finding their more complete senses and usage rules, eventually also their possible closest rendering.

11.4. Idiomatic vulgar

(15) My friends came into my dorm room and distracted me, forced me to drink certain beverages, it ended up not getting done. Oh, and this happened 3 times within the first 2 months of the semester. Needless to say, *I needed to get my shit together*.

Online materials

have/get your shit together Cambridge Advanced Learner's Dictionary & Thesaurus © COonly

have/get your shit together

mainly US offensive

to be or become effective, organized, and skilful:

One of these days I'll get my shit together.

SMART Vocabulary: related words and phrases

Managing and organizing: administer - administration - anti-bureaucracy - bronze command - bureaucracy - conduct - coordinate - fix - have sth in hand idiom - head sth up - in hand idiom - kaizen - management - organization - rationalization - reorganization - reorganize - restructure - restructuring - shit

(16) <https://www.merriam-webster.com/dictionary/get%20one%27s%20shit%20together>

Definition of *get one's shit together*

1 *US, vulgar:* to begin to proceed in an effective way : to become prepared, organized, etc. The team's defense just couldn't seem to *get its shit together*.

2 *US, vulgar:* to begin to live one's life in a responsible and mature way I really need to *get my shit together* and start looking for a job.

The students – in their majority – used the Polish equivalent *zebrać się do kupy* lit. ‘to put oneself together’ – in a similar, although seemingly less vulgar, sense.

12. Conclusions

Over the course of the cooperation students gain increasing responsibility and competence to look for additional data, explications, definitions, examples in the web-based materials and the corpus data available there. The Polish students – in the course of such training – become more independent, to a large extent, from their native speaking peers. At the same time, peer-mentoring relationships are also successful as regards English idiom, smooth expression and naturalness in the sense interpreted above. In this way students develop more persuasive foreign language competence, particularly in terms of the use of more colloquial idiom, and recognize the usage contexts of less official style and native ways of expressing intentions.

And yet, there remain differences, depending both on the native – non-native language use as well as on others – more

culture-specific ones. While the American students typically use friendly comments for corrections, the Polish students' remarks convey suggestions by reference to authority (dictionaries/grammar books) rather than to their addressees' native language and their own acquired competences. This distinction can also be attributed to a higher expressivity of American vis-a-vis Polish culture, although, both would be rather considered more expressive when contrasted with the reserved ones such as the British cultural behaviour patterns (Lewandowska-Tomaszczyk and Wilson 2021) The present study shows to what extent Polish students' exposure to more natural academic and colloquial styles of American English, and the relations they developed with the American peers, can be linked to a more thorough study and reflection on the materials and use of samples of the language in online dictionaries, authentic American English materials, digital data and corpus materials. As shown in the post-task reflective memos (Lewandowska-Tomaszczyk and Veeramoothoo 2021) such procedures and tasks make the Polish students more proficient in English and contribute to the raising of their translation skills and to their digital skills development, while both groups – more sensitive to the intercultural and cross-linguistic relations.

References

- Allard, Carson C., Patricia F. Goldblatt, Jacqueline I. Kembball, Shirley A. Kendrick, K. Judith Millen, Déirdre M. Smith. (2007). "Becoming a reflective community of practice". *Reflective Practice* 8/3: 299-314.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. London/New York: Longman.
- Calderhead, James (1987). "Reflective teaching and teacher education". *Teaching and Teacher Education* 5/1: 43-51.
- Farr, Fiona, Elaine Riordan (2012). "Students' engagement in reflective tasks: an investigation of interactive and non-interactive discourse corpora". *Classroom Discourse* 2/3: 126-143.

- Hofstede, Geert (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.
- Humbley, John, Bruce Maylath, Birthe Mousten, Sonia Vandepitte, Lucy Veisblat (2005). "Learning localization through Trans-Atlantic collaboration". In: George F. Hayhoe (ed.) *Proceedings of the IEEE International Professional Communication Conference, 10-13 July 2005, University of Limerick, Ireland*. New York: IEEE, 578-595.
- Geeraerts, Dirk (2015). "Quantitative corpus onomasiology. Conceptual change" Digital Humanities Case Studies Lectures given on 7 Dec – 8 Dec 2015. KU Leuven.
- Goldberg, Adele (1995). *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Lakoff, George (1987). *Women, Fire and Dangerous Things: What Categories Reveal About Human Mind*. Chicago: University of Chicago Press.
- Langacker, Ronald W. (1987-1991). *Foundations of Cognitive Grammar*. Vols. 1-2. Stanford: Stanford University Press.
- Langacker, Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Lewandowska-Tomaszczyk, Barbara (1989). "Praising and complimenting". In: Wiesław Oleksy (ed.). *Contrastive Pragmatics*. Amsterdam: Benjamins. 73-100.
- Lewandowska-Tomaszczyk, Barbara (2001). "Dictionaries, Language Corpora and Naturalness in Translation". In: Marcel Thelen, Barbara Lewandowska-Tomaszczyk (eds.). *Translation and Meaning*. Part 5. Maastricht: Universitaire Pers Maastricht. 177-186.
- Lewandowska-Tomaszczyk, Barbara (2004). *Gramatyka angielska na autentycznym materiale korpusowym*. Łódź: Wyższa Szkoła Studiów Międzynarodowych w Łodzi.
- Lewandowska-Tomaszczyk, Barbara (2007). "Polysemy, prototypes, and radial categories". In: Dirk Geeraerts, Hubert Cuyckens (eds.). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 139-169.
- Lewandowska-Tomaszczyk, Barbara (2010). "Re-conceptualization and the emergence of discourse meaning as a theory of translation". In: Barbara Lewandowska-Tomaszczyk, Marcel Thelen (eds.). *Meaning in Translation*. Frankfurt: Peter Lang, 105-148.
- Lewandowska-Tomaszczyk, Barbara (2012a). "Explicit and tacit: An interplay of the quantitative and qualitative approaches to translation". In: Michael P Oakes, Ji Men (eds.). *Quantitative Methods in Corpus-based Translation Studies: A practical guide to descriptive*

- translation research*. Amsterdam: John Benjamins Publishing, 1-34.
- Lewandowska-Tomaszczyk, Barbara (2012). "Cognitive corpus analysis – A new agenda for Contrastive Studies". *Connexion – A Journal of Humanities and Social Sciences*, 29–63.
- Lewandowska-Tomaszczyk, Barbara (2017). "Cluster equivalence, general language, and language for specific purposes". In: Marcin Grygiel (ed.). *Cognitive Approaches to Specialist Language*. Newcastle upon Tyne: Cambridge Scholars Publishing, 384-418.
- Lewandowska-Tomaszczyk, Barbara (2021). "Comparing languages and cultures: Parametrization of analytic criteria (Сравнение языков и культур: параметризация аналитических критериев)." *Russian Journal of Linguistics* 25/2: 343–368.
- Lewandowska-Tomaszczyk, Barbara (2021a). "Judgment and argumentation in cross-cultural online collaboration between native and non-native students." In: Anna Turula, Anita Buczek-Zawiła (eds.). *CALL for Background*. Berlin: Peter Lang, 33-49.
- Lewandowska-Tomaszczyk, Barbara (2022). "Meaning". In: *ENTI – Encyclopedia of Translation and Interpreting*. ISSN: 2951-6714. <https://www.aieti.eu/en/encyclopaedia/home/>
- Lewandowska-Tomaszczyk, Barbara, Łukasz Bogucki (2016). "Volunteer translation, collaborative knowledge acquisition and what is likely to follow." In: Łukasz Bogucki, Barbara Lewandowska-Tomaszczyk, Marcel Thelen (eds.). *Translation and Meaning. New Series*. Vol. 2. Pt 2. Frankfurt a. Main: Peter Lang, 37-46.
- Lewandowska-Tomaszczyk, Barbara, Piotr Pezik (2018). "Parallel and comparable corpora, cluster equivalence and translator education." In: *Society and Languages in the Third Millennium – Communication. Education. Translation*. RUDN University: Moscow, 131-142.
- Lewandowska-Tomaszczyk, Barbara, Heather Slomski (2016). "Collaboration in language development between American and Polish university students". *Konin Language Studies*. Vol. 4. *Collaborative Practices in Language* 3: 305-330.
- Lewandowska-Tomaszczyk, Barbara, Saveena Veeramoothoo. (2021). "Language and technology skills development in cross-cultural virtual teams". In: Elisabet Arnó Macià, Marta Pérez Aguilar, Judith Borràs, Comas Mancho, Balbina Moncada, Dietmar Tatzl (eds.). *Multilingual academic and professional communication in a networked world. Proc. AELFE-TAPP 2021 (19th AELFE conference, 2nd TAPP conference)*. Universitat Politècnica de Catalunya, 1-9.

- Lewandowska-Tomaszczyk, Barbara, Paul A. Wilson. (2021). "Expressive and Reserved Cultural Linguistic Schemas: British and American Pride Clusters". In: Marzieh Sadeghpour, Farzad Sharifian (eds.). *Cultural Linguistics and World Englishes*, 261-293.
- Monteiro, Kátia (2014). "Corrective feedback during video-conferencing." *Language Learning and Technology* 18/3: 56-79.
- Neubert, Albrecht (2000). "Competence in Language, in Languages, and in Translation". In: Christina Schaeffner, Beverly Adab (eds.). *Developing Translation Competence*. Benjamins Translation Library 38. Amsterdam: Benjamins. 3-18.
- Nora, Pierre (1992). *Les Lieux de Memoire*. Paris: Gallimard.
- Panova, Iliana, Roy Lyster (2002). "Patterns of Corrective Feedback and Uptake in an Adult ESL Classroom". *TESOL Quarterly* 36/4: 573-595.
- Peżik, Piotr (2012). "Wyszukiwarka PELCRA dla danych NKJP". In: Adam Przepiórkowski, Mirosław Bańko, Rafał Górski, Barbara Lewandowska-Tomaszczyk (eds.). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN, 253-279.
- Peżik, Piotr (2014). "Graph-based analysis of collocational profiles". In: Vida Jesenšek, Petr Grzybek (eds.). *Phraseologie im Wörterbuch und Korpus*. Praha: ZORA 97 Filozofska fakuteta, 227-243.
- Sharifian, Farzad (2017). *Cultural Linguistics*. Amsterdam: Benjamins.
- Sinclair, John (1983). "Naturalness in Language". *Ilha do Desterro: A Journal of English Language Literatures in English and Cultural Studies* 2/11: 45-55.
- Snell-Hornby, Mary (2006). *The Turns of Translation Studies: New Paradigms or Shifting Viewpoints?* Amsterdam/Philadelphia: John Benjamins.
- Surowiecki, James (2005). *The Wisdom of Crowds*. New York: Anchor Books, Division of Random House.
- Tirado, R. Guzmán (2019). "Translation of words with a cultural component (Based on the Spanish translation of the novel by Eugene Vodolazkin "The Aviator")". *Russian Journal of Linguistics* 23/2: 473-486.
- Trompenaars, Fons. Charles Hampden Turner. (1997). *Riding the Waves of Culture: Understanding Cultural Diversity in Business*. 2nd ed. London – Santa Rosa: Nicholas Brealey Publishing Limited.

Corpora, online dictionaries and references materials

British National Corpus (BNC) <http://www.natcorp.ox.ac.uk/>
National Corpus of Polish (nkjp.pl) Narodowy Korpus Języka Polskiego
(NKJP) \ <http://nkjp.uni.lodz.pl/>
PELCRA_CLARIN Tools <http://pelcra.clarin-pl.eu/>
paralela <http://paralela.clarin-pl.eu/>
<http://paralela.clarin-pl.eu/#search/pl/-1/lowest%20bidder/-1/0/20/0/true/0/true/-1/-1/-1/-1/source>
Corpus of Contemporary American English <https://www.english-corpora.org/coca/>
NOW corpus <https://www.english-corpora.org/now/>
TAPP <https://sites.psu.edu/massimoverzella74/what-is-tapp/>
Online Dictionaries @ corpora
<https://www.quora.com/What%E2%80%99s-a-%E2%80%9Cmilitary-grade-weapon%E2%80%9D>
<https://www.imdb.com/title/tt3507312/>
<https://www.merriam-webster.com/dictionary/get%20one%27s%20shit%20together>
eur-lex.europa.eu

Barbara Lewandowska-Tomaszczyk
ORCID iD: 0000-0002-6836-3321
Department of Language and Communication
University of Applied Sciences in Konin
Przyjaźni 1
62510 Konin
Poland
barbara.lewandowska-tomaszczyk@konin.edu.pl

Using the LARA platform to crowdsource a multilingual, multimodal Little Prince

ELHAM AKHLAGHI, ANNA BĄCZKOWSKA,
BRANISLAV BĚDI, HAKEEM BEEDAR, CATHY CHUA,
CATIA CUCCHIARINI, HANIEH HABIBI,
IVANA HORVÁTHOVÁ, CHRISTÈLE MAIZONNIAUX,
NEASA NÍ CHIARÁIN, NATHALIE PATERSON,
CHADI RAHEB, MANNY RAYNER, CHUNLIN YAO

*Received 28.02.2022,
received in revised form 24.11.2022,
accepted 25.11.2022.*

Abstract

We describe an ongoing project, in which an informally organised international consortium is using the open source LARA platform to create multimodal annotated editions of Antoine de Saint-Exupéry's *Le petit prince* in multiple languages, so far French, English, Italian, Icelandic, Irish, Japanese, Polish, Farsi and Mandarin. LARA versions of the book include integrated audio and translations and an automatically generated lemma-based concordance, and are freely available online. We describe the methods used to construct the various versions. In some cases, work for a given language was simply divided by type, typically with one person adding translations and another recording audio. In other languages, we experimented with crowdsourcing methods, splitting the text into chapter-sized units and using the LARA platform to distribute these to multiple annotators, then com-

bining the results at the end. Finally, we report an initial classroom study, where the French version was used by intermediate-level Australian students of French.

Keywords

Computer Assisted Language Learning, multimedia, crowdsourcing, English, Farsi, French, Icelandic, Irish, Italian, Japanese, Mandarin, Polish

Wykorzystanie crowdsourcingu w wielojęzycznym i multimodalnym opracowaniu „Małego Księcia” na platformie LARA

Abstrakt

Artykuł opisuje projekt, w którym nieformalnie zorganizowane międzynarodowe konsorcjum wykorzystuje platformę open source LARA do stworzenia multimodalnych, anotowanych tłumaczeń „Le petit prince” Antoine'a de Saint-Exupéry'ego. Tłumaczenia te obejmują kilka języków, w tym język francuski, angielski, włoski, islandzki, irlandzki, japoński, polski, farsi i mandaryński. Na platformie LARA dostępne są przekłady ww. książki, zintegrowane z nimi nagrania audio, a także automatycznie generowane konkordancje oparte na lematach. W artykule opisane zostały metody użyte do skonstruowania poszczególnych wersji językowych. Praca nad tworzeniem przekładów organizowana była w dwojaki sposób. Dla niektórych języków zadania dzielono na dwie osoby, z których jedna dodawała tłumaczenia, a druga nagrywała do nich dźwięk. W innych przypadkach eksperymentowano z metodami crowdsourcingowymi, dzieląc tekst na rozdziały i wykorzystując platformę LARA do przydzielania fragmentów tekstu wielu anotatorom, a następnie łącząc wyniki ich pracy w jeden tekst. Ponadto, w artykule opisano wstępne badania dotyczące wykorzystania platformy LARA w kontekście edukacyjnym, gdzie australijscy studenci na średnim poziomie zaawansowania językowego korzystali z przekładu „Le petit prince” w toku nauki języka francuskiego.

Słowa kluczowe

Computer Assisted Language Learning, multimedia, crowdsourcing, j. angielski, j. perski, j.francuski, j. islandzki, j. irlandzki, j.włoski, j. japoński, j. mandaryński, j. polski

1. Introduction

In this paper, we present a case study using the Learning And Reading Assistant (LARA; Akhlaghi et al 2019; <https://www.unige.ch/collector/lara/>), an open platform that has been under development by an international consortium since 2018. The purpose of LARA is to enable easy conversion of plain texts into a multimedia form designed to support development of reading skills, adding embedded audio, translations and other annotations, and typically addresses high beginner and intermediate level learners. Related functionality is offered by several other currently available platforms, high-profile examples including LingQ¹ and LearningWithTexts².

The overall goal of the project we describe here is to create LARA editions, in many languages, of Antoine de Saint-Exupéry's classic short novel *Le petit prince* ("The Little Prince"). We selected this text for several reasons. The book is out of copyright, has been translated into nearly four hundred languages, and is widely appreciated by both children and adults. The grammar and vocabulary are easy enough to be approachable at an intermediate level, but hard enough to stretch the reader; similarly, the length, about 15,000 words, is short enough to be manageable but long enough to give a sense of achievement if the student can reach the end. These positive factors combine to make the book hugely popular as an intermediate level reader.

1 <https://www.lingq.com/>

2 <https://sourceforge.net/projects/lwt/>

◀ ▲ ▶ Table of Contents (1)

▶ 0:00 / 8:54 (2)

CHAPITRE X (3) (4)

(5)
Il se trouvait dans la région des astéroïdes 325, 326, 327, 328, 329 et 330. (6) Il commença donc par les visiter pour y chercher une occupation et pour s'instruire. (6)

La première était habitée par un roi. (6) Le roi siégeait, habillé de pourpre et d'hermine, sur un trône très simple et cependant majestueux. (6)

(6)
se trouver

(7)
← N'oubliez pas que je **me trouvais** à mille milles de toute région habitée. (6) \

← Il **se trouvait** dans la région des astéroïdes 325, 326, 327, 328, 329 et 330. (6) \

← - C'est un savant qui connaît où **se trouvent** les mers, les fleuves, les villes, les montagnes et les déserts. (6) \

← Mon étoile **se trouvera** juste au-dessus de l'endroit où je suis tombé l'année dernière... (6) \

← Si tu aimes une fleur qui **se trouve** dans une étoile, c'est doux, la nuit, de regarder le ciel. (6) \

← C'est trop petit chez moi pour que je te montre où **se trouve** la mienne. (6) \

[Frequency index](#)

[Alphabetical index](#)

Figure 1

A sample page from the French LARA edition of *Le petit prince*. The text is in the upper part of the screen. Controls at the top (1) support navigation; each page contains a control (2) that plays audio for the complete page. Audio for a sentence can be played using the loud-speaker icons (3), and hovering over a pencil icon (4) shows a pop-up translation for the sentence. Clicking on a word plays audio for it and also shows a concordance; here, the user has just clicked on the multiword *se trouvait* (“found himself”) (5), bringing up the concordance for the lemma *se trouver* (“to find oneself”) (6). Clicking on a backarrow in the concordance (7) moves to the place in the main text where that sentence occurs.

The specific languages we have used are English, Farsi, French, Icelandic, Italian, Irish, Japanese, Mandarin and Polish: in some cases, work is complete, in others it is still in progress. We have experimented with a number of methods for dividing up the work required to create the LARA versions. For English, French, Icelandic, Italian, Irish and Mandarin, we simply assigned responsibility for the whole of each type of work to a single person. Thus, for example, one person added the translations for that language, one recorded the embedded audio, and so on. In Polish, Farsi and Japanese, we crowdsourced at the level of the text; we divided it into chapter-length pieces and assigned each piece to a different person, combining the results at the end. For Polish, this was done manually. For Farsi and Japanese, we have gone further and added infrastructure to the LARA platform to make the process automatic, with the platform itself taking responsibility for distributing the pieces of text and recombining the annotations added by the crowd-workers.

The rest of the paper is organised as follows. Section 2 presents background on the LARA platform, and Section 3 describes the new LARA crowdsourcing component. Section 4 describes the work we carried out to build the different versions of *The Little Prince*; in several cases (Farsi, Italian, Japanese, Mandarin and Polish), this involved extending the platform to provide better support for the language in question. Section 5 presents an initial study where the French version was trialed in the context of an intermediate-level course. The final section concludes and outlines further directions.

2. The LARA platform

Learning And Reading Assistant (LARA; <https://www.unige.ch/collector/lara/>; Akhlaghi et al 2019) is an open source project whose goal is to develop easily accessible tools that support conversion of texts into an annotated multimedia form designed to support non-native learners. LARA has its roots in enetCollect (Lyding et al 2019; <https://enetcollect.eurac.edu>), a European COST network that links together about 200 researchers intere-

sted in the intersection of crowdsourcing and CALL. Initial funding came from a Swiss National Science Foundation grant to the University of Geneva, under the SNSF's COST program. The project has since expanded to involve individuals and groups in over a dozen countries. After Switzerland, still the de facto centre, the most active groups are the ones in Iceland, Iran, Ireland, Australia, Holland and Poland. Publications and links to examples of LARA content can be found on the LARA site.

The LARA software platform consists of three levels. The lowest level, the *core engine*, was the first to be developed; it consists of a set of Python scripts which carry out the central functions of converting text and other resources (recorded audio, translations, etc) into multimedia form. It is possible to use the core engine on its own by invoking it from the command line. In practice, however, most people access LARA through the *portal* (<https://lara-portal.unige.ch/>) which provides an easy-to-use wizard style interface, implemented in PHP. The third layer, in development, is the *social network layer*. As the name suggests, this will provide a top level which organises LARA users and content in a social network inspired way, with home pages for each user and piece of content and the expected links between them. In this paper, since we are focusing on the content creation process, we will mostly consider relevant parts of LARA's portal level functionality.

Figure 1 above shows an example of LARA content, a page taken from the French edition of *Le petit prince*, and illustrates the main types of annotations supported by LARA: audio, translations and multi-words. We now briefly describe how these annotations are created. Full details can be found in the online documentation (Rayner et al, 2021).

2.1. Portal top level

The portal top level provides easy access to the other types of functionality offered. The greater part of this involves the content creation and editing process, relevant aspects of which are described in sections 2.2-2.5. §3 describes the new crowdsour-

cing mechanism, which is being used for the first time in the project presented here.

2.2. Lemma tagging

The first step in the annotation process is to process the text so as to tag each word with its associated lemma. This can be done by hand, but for most of the languages we work with³ is in practice performed by a tagger/lemmatiser, followed by manual post-editing of the automatically added tags. The LARA portal has several tagger/lemmatisers integrated so that the initial automatic tagging step can be invoked with a single click. Our experience shows that the performance of modern tagger/lemmatiser packages for this kind of task varies a great deal. In good cases (well-resourced languages like English and French; texts similar to those used to train the resources) error rates are in low single digits, and an experienced editor can carry out the post-editing task at a rate of several thousand words per hour (Akhlaghi et al, 2020). In less favorable cases, some of which are described in this paper, error rates can be a good deal higher. An example of tagged LARA text is shown in Figure 2.

```
CHAPITRE X||
Il se trouvait#trouver# dans la#le# région des#du# asté-
roïdes#astéroïde# 325, 326, 327, 328, 329 et 330.|| Il
commença#commencer# donc par les#le# visiter pour y cher-
cher une#un# occupation et pour s'#se#instruire.||
La#le# première#premier# était#être# habitée#habiter# par
un roi.|| Le roi siégeait#siéger#, habillé#habiller# de
pourpre et d'#de#hermine, sur un trône très simple et
cependant majestueux.||
```

Figure 2

Example of tagged LARA text automatically produced by an integrated tagger/lemmatiser. Lemma tags are delimited with # ... #, segments with ||, compound words with |.

³ LARA documents have been created for some languages where no tagger/lemmatiser exists.

2.3. Audio recording

Audio recording is performed using the LiteDevTools online recording tool (LDT; Akhlaghi et al, 2019). Once the lemma tagging step is complete, the LARA portal automatically uploads recording scripts for segments and words to LDT, where they can be accessed by the voice talent. The simple and intuitive interface is shown in Figure 3.

Audio can also be created using an integrated TTS engine; the portal currently supports ReadSpeaker, Google TTS and ABAIR for Irish (ABAIR 2021). Comparison between human and TTS audio suggests that the gap between them is now quite narrow. In a recent empirical study we carried out, subjects preferred TTS to some of the human voices used (Akhlaghi et al 2021).

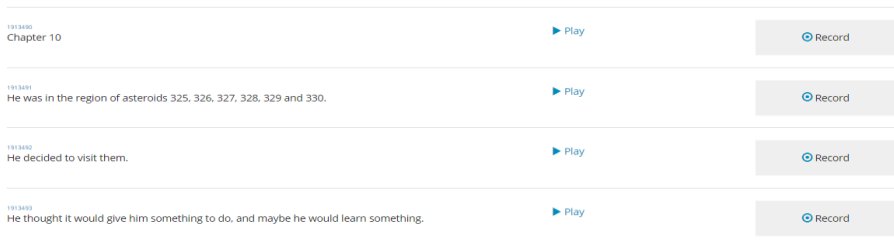


Figure 3

Recording audio with the online LDT tool. The voice talent clicks 'Record', reads the text on the left, and clicks again to stop. They can listen to the recorded audio using the 'Play' button and, if necessary, rerecord it. Items can be attempted in any order, and the voice talent is free to log out and return later. If the source text is edited in the LARA portal, the recording tasks are automatically updated to reflect this.

2.4. Translations

Sentence-level (more properly, segment-level) translations are entered using another simple portal interface, illustrated in Figure 4.

CHAPITRE X
Chapter 10
Il se trouvait dans la région des astéroïdes 325, 326, 327, 328, 329 et 330.
He was in the region of asteroids 325, 326, 327, 328, 329 and 330.
Il commença donc par les visiter pour y chercher une occupation et pour s'instruire.
He decided to visit them. He thought it would give him something to do, and maybe he would learn something.

Figure 4

Segment-level translation interface. The portal presents the segments in the reading language, the annotator fills in the translations in the student language. Segments are presented in the order in which they occur in the text.

Word translations are created in two stages. First, the annotator fills in translations for word types, using an interface similar to the one used for segments. This is shown in Figure 5. Word type translations are shared between different texts as a common resource.

french	english
bonjour	hello
renard	fox
être	be
fois	time
jour	day
dessin	drawing

Figure 5

Word type translation interface. The reading language words are presented on the left, the annotator fills in the student language words on the right. Hovering over one of the words on the left shows a pop-up with a list of occurrences of the word in the current text.

The word type translations are then used to populate a set of pages (“word token translations”) which allow the annotator to change the translation of a word in a specific context. An example of a word token translation page is shown in Figure 6.

CHAPITRE	X														
Chapter	X														
Chapter	10														
Il	se	trouvait	dans	la	région	des	astéroïdes	325	326	327	328	329	et	330	
He	himself	found	in	the	region	of the	asteroids	325	326	327	328	329	and	330	
He	was	in	the	region	of	asteroids	325	326	327	328	329	and	330		
Il	commença	donc	par	les	visiter	pour	y	chercher	une	occupation	et	pour	s'	instruire	
He	began	so	by	them	visiting	to	there	look for	an	occupation	and	to	him	instruct	
He	decided	to	visit	them	He	thought	it	would	give	him	something	to	do	and	maybe he would learn something

Figure 6

Word token translation interface. Lines are presented in groups of three. The first line is a sentence from the text, the third is the sentence translation, and the annotator fills in the word glosses in the middle. Initial values are set from the type translations. Multiword expressions are highlighted in red.

2.5. Multi-word expressions

Multi-word expressions (MWEs) are common in most languages. We will use this term in a broad sense, to refer to any sequence of words, for example a phrasal verb, which intuitively should be treated as a unit. An occurrence of an MWE often, though by no means always, consists of a contiguous set of words. Thus, for example, the continuous sequence “just about” in “I had **just about** enough water for eight days” is treated as an MWE, and similarly the discontinuous sequence “looked ... like” in “They **looked** exactly **like** his own flower” is treated as a form of the MWE “look like”.

LARA provides a systematic mechanism for annotating MWEs. Each language includes an MWE lexicon, which in general has been compiled partly from LARA texts and partly from

sources on the web. An entry in the MWE lexicon is a list of words, each of which can be either a surface word (the word must appear exactly as listed in the MWE entry), or a lemma (the word has the named lemma tag). In the notation used, lemmas are distinguished by adding asterisks around them. Thus, in the two examples above, the lexicon entry for “just about” is

```
just about
```

while the lexicon entry for “look like” is

```
*look* like
```

Other features make it possible to define lexicon entries for MWEs like reflexive verbs, by introducing the possibility of defining a lexicon token to represent a class of words. Thus, for example, in the French MWE lexicon, the token *se* is defined to represent the class of reflexive pronouns, which means that a lexicon entry like

```
se *coucher*
```

(“oneself lie-down”, “go to bed”) matches any of the sequences *me couche* (1st person singular), *vous couchez* (2nd person plural), etc. A full description of the MWE lexicon notation can be found in the online documentation but is somewhat beside the point here, since the annotator does not see the lexicon infrastructure. All they see is a list of possible MWE occurrences, presented using the interface shown in Figure 7. Their task is to mark each occurrence as genuine or spurious.

<p style="text-align: center;">ressembler à in:</p> <p>Mais celle-là avait germé un jour, d'une graine apportée d'on ne sait où, et le petit prince avait surveillé de très près cette brindille qui ne ressemblait pas aux autres brindilles.</p>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<p style="text-align: center;">se montrer in:</p> <p>Et puis voici qu'un matin, justement à l'heure du lever du soleil, elle s'était montrée.</p>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<p style="text-align: center;">il y avoir in:</p> <p>- Il n'y a pas de tigres sur ma planète, avait objecté le petit prince, et puis les tigres ne mangent pas l'herbe.</p>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<p style="text-align: center;">se mettre in:</p> <p>- Le soir vous me mettez sous globe.</p>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Figure 7

Interface for marking occurrences of multiword expressions. The candidates are ordered by the number of words skipped, with contiguous occurrences first. In the screenshot above, the first three candidates are marked as genuine occurrences of MWEs, the fourth as spurious.

3. Adding support for crowdsourcing in LARA

The functionality offered by the LARA crowdsourcing interface is straightforward and intuitive. When creating a LARA project, the developer may optionally divide the text into a number of sections by inserting separator tags. If they have done this, they can then use a dashboard control to *crowdsource* the project. The effect is that a set of subprojects are created, one for each section. Each subproject has the same settings as the original project, except that it contains only the relevant portion of the corpus.

The crowdsourced subprojects are made visible under the *Available Tasks* tab of the crowdsourcing dashboard, which is accessible from the portal top level. Any user who wishes can *pick up* a subproject from the Available Tasks tab. This transfers the subproject to its new owner, who is then free to add any of the types of annotations described under §§2.2-2.5 above. However, the owner of the original project (the *task requester*) can still view the crowdsourced subtask. They can also, at any time, use another dashboard control to *collect* the current versions of

the subprojects. As the name suggests, this means that the texts and annotations currently present in the subprojects are combined back into the parent project, so that the task requester can obtain an up-to-date snapshot of the project as a whole. The task requester and the crowdworker can communicate through the crowdsourcing dashboard by updating the message thread attached to each subproject. It is possible for the crowdworker to *give back* a task they no longer wish to do, and for the task requester to *take back* a task if they feel insufficient progress is being made.

A detailed description of the crowdsourcing functionality can be found in the online documentation⁴.

Table 1

Status of “LARA Little Prince” projects by chapter for the nine languages currently finished or under development, as of May 18 2021.

Languages are ordered by completeness.

Language	Chs	Tag- ging cor- rected	Audio		Translations	
			Sents	Words	Sents	Words
French	27	All	All (Human)	All (TTS)	All	All
English	27	All	All (Human)	All (TTS)	All	All
Italian	27	All	All (Human)	All (Human)	All	No
Farsi	27	All	26 chs (Human)	All (Human and TTS)	25 chs	22 chs
Icelandic	27	All	All (TTS)	All (TTS)	No	No
Polish	18	18 chs	18 chs (Human)	18 chs (Human)	18 chs	18 chs

⁴ <https://www.issco.unige.ch/en/research/projects/collector/LARADoc/build/html/crowdsourcing.html>

Japanese	27	3 chs	All (TTS)	All (TTS)	9 chs	1 ch
Irish	2	2 chs	2 chs (TTS)	2 chs (TTS)	2 chs	2 chs
Mandarin	1	1 ch	1 ch (Human)	1 ch (Human)	1 ch	1 ch

4. Creating LARA versions of *The Little Prince* in different languages

In this section, we describe how we created LARA versions of *The Little Prince* in English, Farsi, French, Icelandic, Italian, Irish, Japanese, Mandarin and Polish. The work involved was of two kinds. In all cases, texts were annotated using the methods described in Sections 2 and 3 to add embedded audio and translations, and tag occurrences of multiword expressions. For some languages, we first needed to extend the platform's support for the language in question, most often by adding or improving automatic tagging capabilities. Links to the LARA versions, in many cases still ongoing, are posted on the LARA examples page⁵; a summary of the current state of play, as of May 2022 is presented in Table 1 above. We group the languages by the type of work involved.

4.1. French, English, Icelandic and Irish

These four languages have been used extensively since the start of the LARA project, and no further work was required on infrastructure. English and French perform tagging and lemmatization using TreeTagger with the appropriate packages (Schmidt 1994); Icelandic uses a combination of the Greynir, Nefnir and ABLtagger packages (Steingrímsson 2019, Akhlaghi et al 2020). In all four languages, work was organised so that one person was responsible for post-editing tagging, adding translations, and marking MWEs. For French and English, audio was recor-

⁵ <https://www.unige.ch/collector/lara-content>

ded by a child/young teen (12-year-old boy for French; 15-year-old girl for English) paid at the local babysitter rate. For Icelandic and Irish, we used TTS engines, ReadSpeaker for Icelandic and ABAIR for Irish.

4.2. Italian

For Italian, we also used TreeTagger for tagging and lemmatization, but some work was needed to handle postverbal clitics. Italian orthography affixes these directly to the verb (*masticarla* = *masticar-la* = “chew it”; *spiegarglielo* = *spiegarglie-le* = “explain them to him”). However, we found that TreeTagger only tags words of this kind with the lemma for the verb, discarding the clitics. We corrected by adding post-processing code which operates on words that have been tagged as verbs of the relevant types and end in affixes which match legal sequences of postverbal clitics.

This time, annotation work was divided up so that correction of tagging, and marking of MWEs, was performed by two people, one native and one near-native in Italian, while a third person added translations. As with French and English, audio was recorded by a young teen voice, this time a 14-year-old girl.

4.3. Polish

When we started work on the Polish edition, we soon determined that performance for the TreeTagger Polish package was very poor. As this is also the case for the Russian and Slovak packages, it may be that the TreeTagger architecture is not well suited to the highly inflected Slavic languages. We looked around for alternatives, and found that the combination of the Morfeusz2 lemmatiser (Woliński 2017) and Concraft tagger was recommended as giving state-of-the-art performance. As in the case of the Italian TreeTagger package, integrating the Morfeusz2/Concraft combination was not entirely trivial. For complex reasons having to do with the grammar of Polish, past tense verb affixes are analysed by Morfeusz2 as separate morphemes,

but this approach is not appropriate for LARA, which needs to analyse a verb as a single word. After some discussion with the Morfeusz team, who were helpful in explaining the issues, we found that we could solve the problem in post-processing by joining together verbs and their associated past tense morphemes.

Departing from the recipe we had used for the five languages above, we experimented with a new one for Polish. The text was manually divided into chapters and distributed among the members of a course which the Polish member of the consortium was teaching. Each student was responsible for all aspects of annotation for their chapter. At the end, a script was used to recombine the annotations from the multiple users into a single LARA text, following which we did a little post-editing to add basic markup for headings, images and division into pages.

This experiment was not entirely successful as there were only enough students to assign 20 out of 27 chapters, and some students did not adequately complete their assignments. The majority, however, did their tasks well, giving us valuable information to suggest that the crowdsourcing method was a step in the right direction. This motivated the development of the crowdsourcing infrastructure previously described in §3.

4.4. Farsi

Farsi (Persian) is an Indo-European language written with an Arabic script. We perform tagging and lemmatization using the hazm package.⁶ A noteworthy property of Farsi is the unusually high frequency of phrasal verbs. This meant that an early priority in developing Farsi resources for LARA was to compile a substantial lexicon of phrasal verbs compatible with the LARA MWE processing module (cf. §2.5).

The Farsi version was produced using the automatic crowdsourcing method described in §3. The tagging phase was first carried out by one of the authors, a Farsi native speaker with

⁶ <https://pypi.org/project/hazm/0.3/>

long-time experience of using LARA, after which the text was divided into 27 chapters and put on the crowdsourcing platform. Students enrolled in a course taught by the author who performed the tagging were invited to pick up chapters and add human audio and translations. They were told that completion of the assignment within a three-week window would be rewarded with a 10 % bonus credit on the course.

16 students completed their chapter and 10 more returned a partial result. Missing word audio, the most common missing part, was added using the ReadSpeaker integration of the Ariana TTS engine.⁷

This exercise was the first large-scale test of the crowdsourcing mechanism from §3. It performed correctly, with no obvious technical faults, but several of the students complained that the portal was not user-friendly enough for casual users who had no previous experience with it.

4.5. Mandarin

Chinese, the first non-Indo-European language of the ones we used, posed special problems at the level of infrastructure. The highly analytical nature of Chinese grammar means that lemmatization as such is not required. This is replaced by the well-known Chinese segmentation problem. Since standard Chinese orthography does not mark spaces between words, the first task in text processing is typically to segment the input. After experimenting with a couple of alternatives, we decided to use the popular Jieba package⁸, which was easy to integrate. It was also important to provide word glosses in pinyin (Roman alphabet), the pedagogical value of which is well attested. There are many sites on the web which provide accurate pinyin conversion tools. We integrated the one from chineseconverter.com.⁹

⁷ <http://www.farsireader.com/english/>

⁸ <https://github.com/fxsjy/jieba>

⁹ <https://www.chineseconverter.com/en/convert/chinese-to-pinyin>

The above issues make the Chinese workflow slightly different from the one used for the other languages. Instead of correcting an automatically tagged version of the text, the annotator starts by correcting an automatically segmented version; thus they edit segment boundaries rather than lemma tags. Figure 8 illustrates.

As of May 2022, one sample chapter has been produced for Mandarin.

当[我]还[只有]六岁[的]时候[，]在[一本]描写[原始森林]的[名叫]《[真实]的[故事]》的[书]中[，]
看到[了]一副[精彩]的[插画]，画[的]是[一条]蟒蛇[正在]吞食[一只]大[野兽]。||页头[上]就是
[那副画]的[摹本]。||

这本书[中]写道[：]“这些[蟒蛇]把[它们]的[猎获物]不[加]咀嚼[地]囫圇[吞下]，尔后就[不能]
再[动弹]了[；]它们就[在]长长的[六个月]的[睡眠]中[消化]这些[食物]。”||

当时[，]我[对]丛林中[的]奇遇[想得]很多[，]于是[，]我也[用]彩色[铅笔]画[出]了[我的]第一
副[图画]。||我的[第一号]作品[。]||它是[这样]的[：]||

我[把]我的[这副杰作]拿[给]大人[看]，我[问]他们[我的画]是不是[叫]他们[害怕]。||

他们[回答]我[说]：“一顶[帽子]有[什么]可怕[的]？”||

Figure 8

Example of hand-editing Chinese segmentation produced by the Jieba package. As in the case of Figure 2, segment boundaries are marked with double bars, ||, and word boundaries with single bars, |. Red is used to show the changes made by the editor.

4.6. Japanese

Similar to Chinese, Japanese as a non-Indo-European language posed significant issues for LARA. Like Chinese, Japanese is normally written without spaces between words, and segmentation is thus again a substantial problem. Unlike Chinese, however, Japanese is primarily a synthetic language, and its morphosyntax demonstrates inflection of agglutinative roots.

At the moment, LARA performs segmentation, lemmatization and tagging using the Google Cloud Natural Language Morphology and Dependency Trees package for Japanese¹⁰. Unfortunately our experience with the current version of the package on various texts shows a high error rate, typically more than two errors per sentence. Errors are frequent both in segmentation and in lemmatization. As an example of the former, the common word *あいだ* (*aida*, literally “interval” but with the sense of “for”) in the typical context *6か月のあいだ* (*rokkagetsu no aida*, “For six months”) is segmented as two lexemes, *あい* (*ai* “love”) and *だ* (*da*, copula). With regard to lemmatization, verbs are typically not tagged with the dictionary form: thus, for example, *出して* (*dashite*, “leaving”) is tagged as *出し* (*dashi*) instead of the expected *出す* (*dasu*). Some verbs are, however, correctly lemmatised; in particular, the irregular *して* (*shite*, doing) is always lemmatised correctly as *する* (*suru*). We suspect that these errors may be teething problems in a new package, perhaps caused by very insufficient training data. The task is further complicated by the fact that pedagogical considerations do not always coincide with standard Japanese linguistic conventions. For example, while morphosyntactically and phonologically *本には* (*hon-ni-wa*, “in the book”) is realised as a single, inflected unit in Japanese, it is more useful for a learner to see it split up into a content word and two polysemous particles (Masuda 2018: 63–96). For all these reasons, tagged Japanese text so far requires extensive post-editing.

The Japanese version is using the automatic crowdsourcing method described in §3. The workflow is similar to that in the Chinese version, though crowd-workers, for the reasons described above, need to revise both segmentation and tagging. Detailed instructions to the crowd-workers are posted online¹¹. Crowd-workers found it natural to divide up work within the

¹⁰ <https://cloud.google.com/natural-language/docs/morphology>

¹¹ <https://www.issco.unige.ch/en/research/projects/callector/LARAJapaneseLPDoc/build/html/index.html>

chapters, with one crowd-worker correcting the segmentation and tagging, handing it to a second crowd-worker who adds the sentence translations, and then taking it back to add the word translations. The LARA platform's crowdsourcing functionality efficiently supports this flexible division of labor. Unfortunately, however, although all technical aspects appear to work correctly, it has so far proved extremely difficult to attract crowd-workers, and only two chapters have been completed.

5. Using the LARA Little Prince in the classroom

Since July 2020, links to the different versions of the LARA Little Prince have been publicly posted. Although our impression is that they have been well received,¹² we wished to get more formal student feedback. With this in mind, we organised the utilization of the French LARA version of *Le petit prince* as part of a low intermediate level French course held at Flinders University, South Australia, using the French LARA version of *Le petit prince*.

5.1. Context/Background

A set of three sessions was devoted to reading the first three chapters in a class of 16 (11F, 5M) A2-B1 level students of French as a foreign language. The students had completed three semesters of study and had spent around 400 hours learning French. Their majors included international relations, law, psychology and education.

The group was fairly homogeneous, motivated and enthusiastic. They were engaged in their studies and generally applied themselves diligently to assignments outside the classroom. Most of them spoke French spontaneously in class.

¹² In particular, a page for the French LARA edition (<https://www.goodreads.com/book/show/54923317-le-petit-prince>) appears on the popular Goodreads review site. As of November 2022 it has 18 ratings, with an average rating of 4.72.

At this level, students were entitled to 4 hours of weekly lessons, divided into two seminars of two hours each. The first seminar was mostly dedicated to reading and writing, with the introduction of new grammatical points. The second seminar focused on oral comprehension and expression, combining exercises of different types, some using authentic material, and others designed specifically for FFL. Songs and short dictations were also included.

5.2. Process/method

The entire course was based on a textbook (*Entre Nous 2*, Editions Maison des Langues), which provided core material and a point of focus for grammatical and thematic features. Numerous additional documents were offered relating to grammar, linguistic and lexical content, as well as to assist oral and writing skills. In support of reading, the teachers-researchers generally offered short texts to be prepared at home and subsequently discussed in class.

Introducing *The Little Prince* was part of an ongoing process of discovering literary texts; the fact that the book was familiar to many of the students was a facilitating element. Due to time constraints, only the first three chapters were used. The overall aims were to allow students to feel more motivated and more confident in their target language reading skills, help them understand the text, and make them aware of the poetry and the literary effects created by the text.

We expected that using the LARA platform would add value by giving students immediate access to word meanings (both translations and other occurrences of the word in the text, sometimes illustrating contextually different meanings), and also to audio. Audio was made available in human-recorded form for sentences and in TTS voice for words. TTS audio was produced by the ReadSpeaker engine, which at a word level gives quality judged comparable to human audio (Akhlaghi et al 2021).

The two teachers designed a worksheet based on each of the three chapters presented on the LARA platform, whilst maintaining the grammatical points programmed for the semester. Each of the worksheets included comprehension questions on the passage (oral or written) as well as questions relating to grammar or conjugation. The following activities were offered:

1. Activities aimed at working on and checking comprehension

- A. An individual activity consisting of listening and reading simultaneously to be discussed in pairs to enable understanding of the chapter (Chapter 1)
- B. An individual activity consisting of listening and reading simultaneously where each student individually initially answered a series of questions relating to the literal understanding of the passage. This was followed by a consideration at a deeper level, that is, first literal and then literary – of the text and its effects on the student-reader (Chapter 1 and Chapter 3)
- C. A listening and reading activity designed to be done individually with the aim of presenting a brief oral summary of the chapter (Chapter 2 and Chapter 3)
- D. Individual listening / reading activity to test comprehension through a multiple-choice response type activity (Chapter 2)

2. Activities aimed at working on pronunciation/prosody and at expanding the lexicon

- A. Observing selected words with respect to 1) their pronunciation 2) their specific occurrences in the text. Students were also invited to observe the tenses or moods used in the case of verbs (Chapter 2)
- B. Students recorded new vocabulary in their exercise book. (Chapter 2, Chapter 3)
- C. An activity aimed at listening and imitating designed to observe and then reproduce the rhythm of the sentence created by the commas (Chapter 2, first paragraph) and

to observe and reproduce the intonation in the dialogue section (“Mais qu’est-ce que tu fais là?”; Chapter 2)

- D. Pronunciation activities to work on the contrast between nasal sounds such as (ON [ɔ̃] and AN or open O [ɔ] and closed O [o]) vowels. This type of activity involved clicking on a particular word and repeating the sentence aloud (Chapter 3)

3. Activities aimed at raising awareness of the literary nature of the book

- A. Analysis of the gradual unveiling of the Little Prince’s character, both from the textual and visual point of view, as well as a question asking the students to put themselves in the shoes of a child reader discovering the image in question of the Little Prince (Week 2, Chapter 2)
- B. Activities based on supporting documents about the text and the author. These documents comprised an audio biography from the TV5 Monde platform, concise iconographic presentations of the text and the author’s life, an article about the Antoine de Saint-Exupéry Foundation, a tribute to the author by the French Air Force, and an interview with the son of Léon Werth, the person to whom the book is dedicated.¹³

5.3. Questionnaire and responses

In the week immediately following the conclusion of the experiment, the students were asked to fill out an English language questionnaire (see the appendix). The ten responses were analysed following Patton (2002). We organise the material under five headings:

¹³ https://www.youtube.com/watch?app=desktop&v=Veq_K08R1H8&ab_channel=QuelleHistoire, <https://ticsenfle.blogspot.com/2014/12/le-petit-prince-ressources-lecture.html>, <https://www.fondationdefrance.org/fr/fondation/fondation-antoine-de-saint-exupery-pour-la-jeunesse>, <https://perelafouine.com/la-patrouille-de-france-rendra-hommage-a-antoine-de-saint-exupery-et-son-petit-prince/>, https://www.youtube.com/watch?v=O5Z-9tyEZ4I&ab_channel=SaintExuperyTV.

1. Previous knowledge of *Le Petit Prince*
2. Practising reading with LARA
3. Practising pronunciation with LARA
4. Work on vocabulary with LARA
5. General comments on the approach and platform, suggestions

Previous knowledge of Le Petit Prince

3 of the 10 students knew the book before the experiment.

Practising reading with LARA

Several learners highlighted how their comprehension was facilitated by the audio and translation annotations. Student feedback in this and following sections is presented as originally provided without correction.

"LARA is a great way to read and at the same time listen to the correct pronunciation. It is great that the meaning of each word can easily be seen (the translation in English; that makes it much easier and more fun as I don't have to look up each unknown word in the dictionary" (student A)

"[reading was] very good. Was helpful to understand pronunciation and learning new vocabulary and sayings" (student C)

"I thought it was very useful because it exposed the class to a classic French story. Reading the text as a class also improved my reading confidence" (student D)

"Listening to the platform, read the text and then reading it through as a class and individual helped my pronunciation and fluency of reading" (student F)

"I thought it was useful that we could hear the words being spoken out loud as we read it. Additionally, the fact that we could hover over a word to discover its meaning in English was extremely useful, and even enabled me to learn new words" (student G)

"It worked well. We were able to effectively cover everything in an easy understandable way" (student H)

"I enjoyed being able to read along with the speaker and being able to find out what the individual words mean. Overall I think it was very successful" (student J)

Practising pronunciation with LARA

The students appreciated being able to listen and repeat the words and phrases, imitating what they heard:

“It was very helpful to listen to the pronunciation of certain words and to repeat it” (student A)

“it was good as we went through the correct pronunciation while we were reading (...) audio was great, LARA platform is really helpful, pronunciation tasks were good (...) the pronunciation is really helpful because you can hear the liaisons and different tenses” (student B)

“Good, the fact that each word is provided with two pronunciations (boy and female) in a sentence but also just the word made it very helpful” (student C)

“studying pronunciation helped me understand the way a word can change in sound with aspects like liaison. Being able to use a platform that also read the words with a click was also very helpful” (student D)

“Being able to isolate individual sentences and words within the text is useful to improve pronunciation and recognition of liaisons between words for pronunciation (...) a strength is (...) being able to easily repeat sections for practise” (student F)

“I thought it was useful that we could listen to each passage sentence or word being read aloud by a French speaker, which helped to properly understand how to pronounce the words I hadn’t previously known” (student G)

“this was a good approach allowing us to see many ways each word might be used, allowing us to properly understand how to pronounce it” (student H)

“Amazing. I had never heard or used LARA before. The use of AI has helped with some pronunciation. I also like the capacity to repeat sections of the text” (student I)

“I thought that it was very effective to be able to hear and repeat the words both on their own and in sentences (...) it helped to be able to listen and go word by word pronunciation wise” (student J)

Work on vocabulary with LARA

Students found the LARA platform helped with the expansion of their vocabulary, many mentioning the large number of new words and idioms they discovered:

“I learned new words and expressions (...) by reading the text, being able to see the pronunciation of unknown words and by answering questions on the worksheet regarding the text and also by looking at single words or expressions (...) I think it is a great way to learn vocabulary. It also helped me write down new words” (student A)

“I thought that this was really helpful as the LARA platform went through all the scenarios/tenses of certain words (...) I thought this went well, however, it would be more helpful if we translated the text more to get a better understanding” (student B)

“Very good. Went over new vocabulary and the platform also tells you the meaning of each word/phrase so you can easily understand” (student C)

“I think I have gained more French vocab since studying the text and also been exposed to more day-to-day French phrases” (student D)

“Studying the vocabulary used in the text and being able to isolate words to see how they are used in various formats is very helpful to build my vocabulary as well as my ability to use it” (student F)

“I thought it was useful to be able to translate each sentence, especially when the sentence was idiomatic” (student G)

“... gave us lots of opportunities to practise and perfect words (...) hovering over a word showed the translation and helped a lot (...) exact translation of paragraphs are sometimes unclear” (student H)

“This was a bit of a ‘curve ball’ for me as I (didn’t) consider it a real learning objective. However, I now appreciate some of its value” (student I)

General comments on the approach and platform, suggestions

There were several positive comments about LARA’s general flexibility/versatility and its ability to combine with relevant pedagogical activities:

“If class missed you can still go over on your own for comprehension and writing” (student C)

“Another strength is that we can access LARA from home to practise” (student D)

“A strength is accessing it from home” (student F)

“Flexibility and ease of repetition outside the class environment has allowed me to listen, read, speak and write more fluently” (student I)

“When reading a book or text where there are several words I don’t know, it can always be a bit difficult to remember all the new vocabulary. Anyways, by answering questions, reading and listening to the text (maybe several times), we really engage with the text and it makes understanding, remembering and learning easier and also interesting” (student A)

“... was very helpful (...) also for forming sentences” (student C)

“I think LARA is a great platform as it is very accessible, simple and provides a lot of helpful tools to break down French vocab, pronunciation and listening (...) I’d like to read more French classics on LARA as I find it very easy to navigate” (student D)

A few comments were, however, more critical of the pedagogical approach, LARA, or both:

“LARA is great the way it is. More translation activities in class would be beneficial for a better understanding” (student B)

“The use of the LARA platform could have been better (...) The approach is good for teaching pronunciation and vocabulary but it is difficult for people who struggle with reading and comprehension (...) I would have preferred if the chapters were read aloud in class rather than having to read them at home” (student E)

“An improvement of the platform could be including the form of grammar when you click on a word” (student F)

“Exact translations of paragraphs are sometimes unclear (...) may be have translations of paragraphs or sentences available” (student H)

“The ‘click’ sound at the end of each sentence can be annoying” (student I)

“Perhaps single subject options (1 or 2 pages) of a journal, contemporary news story etc. that can be played and discussed to create opinions and differing points of view” (student I)

“A weakness is either listening to one sentence or the whole thing and not be able to listen paragraph by paragraph” (student J)

Summary of the questionnaire responses

The high proportion of positive responses suggests the approach using LARA was appreciated by students. In particular, the ability to hover over a word to discover its meaning in English and hear how it is pronounced was evidently considered very useful.

Several students indicated that they would have liked to be able to listen to full paragraphs when working with LARA and that more time needed to be spent in class on translation and ensuring that the content was properly understood by the group as a whole. The general favorable feedback invites continued use of LARA. Next year's course may include either an extended reading of the whole of *Le Petit Prince*, or allow students to discover another French literary work available on LARA.

6. Conclusions and further directions

We have given an overview of the "LARA Little Prince" project, an informal multinational collaboration aimed at producing annotated multimodal editions of *Le petit prince* in many languages using the open source LARA platform. Each language, in general, poses its own special problems; but we have found that these can be solved, and the project is making steady progress. The first few languages are now at the point where the online texts can be used as learning resources. As reported in §5, initial classroom feedback is encouraging.

A novel aspect of the project has been the introduction of a crowdsourcing method to divide up the tasks for a given language between multiple workers. The method was used for Polish and Farsi with partial success. About two-thirds of the chapters were completed in both languages, though the quality was variable. For Japanese, the crowdsourcing method was unfortunately not successful at all. The tasks were posted on an online Japanese forum where one of the authors was an active member, but despite verbal expressions of interest no one picked them up. The sample is small, but one cannot help remarking on the fact that the Polish and Farsi crowd-workers were students who performed the tasks in the context of courses where they would receive extra credit for completing their assignments, while the prospective Japanese crowd-workers had no similar incentive. It seems to be important to think more about issues of motivation and quality control.

We hope to explore these ideas further. People who would like to take part in the project, either in the existing languages or starting new ones, are very welcome to contact us.

Acknowledgements

Work at the University of Geneva was funded by the Swiss National Science Foundation under projects IZCOZ0_177065 and 100015_204503. We gratefully acknowledge ReadSpeaker's help in making their TTS software available to the LARA project.

Appendix: text of the student questionnaire

- 1) Did you read or study *Le Petit Prince* prior to this semester? At school? At home? In French or English?
- 2) What do you think of the way we practised *reading* in the first three chapters of Le Petit Prince class?
- 3) What do you think of the way we studied *pronunciation* in the first three chapters of Le Petit Prince?
- 4) What do you think of the way we worked on *vocabulary* in the first three chapters of Le Petit Prince?
- 5) What were the strengths/the weaknesses of this approach?
- 6) Do you think that using the LARA platform helped you? Why?
- 7) Did you have a look at the LARA recordings in other languages? If yes, which ones? What did you think of them?
- 8) Do you think we made good use of the LARA platform?
- 9) What else could we do or what could we do differently?

References

- ABAIR: An Sintéiseoir Gaeilge – The Irish Language Synthesiser AB AIR (2021). <http://www.abair.ie>. Accessed 20 July 2022.
- Akhlaghi, Elham, Branislav Bédi, Matt Butterweck, Cathy Chua, Johanna Gerlach, Hanieh Habibi, Junta Ikeda, Manny Rayner,

- Sabina Sestigiani, Ghil'ad Zuckermann (2019). "Overview of LARA: A learning and reading assistant". In: *Proceedings of SLATE 2019*, 99-103.
- Akhlaghi, Elham, Branislav Bédi,, Fatih Bektaş, Harald Berthelsen, Matt Butterweck, Cathy Chua, Catia Cucchiarini, Gülşen Eryiğit, Johanna Gerlach, Hanieh Habibi, Neasa Ní Chiaráin, Manny Rayner, Steinþór Steingrímsson, Helmer Strik (2020). "Constructing multimodal language learner texts using LARA: Experiences with nine languages". In: *Proceeding of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 323-331.
- Akhlaghi, Elham, Anna Bączkowska, Harald Berthelsen, Branislav Bédi, Cathy Chua, Catia Cucchiarini, Hanieh Habibi, Ivana Horvátová, Pernille Hvalsøe, Roy Lotz, Christèle Maizonniaux, Neasa Ní Chiaráin, Manny Rayner, Nikos Tsourakis, Chunlin Yao (2021). "Assessing the quality of TTS audio in the LARA learning-by-reading platform". In: Naouel Zoghlami, Cédric Brudermann, Muriel Grosbois, Linda Bradley, Sylvie Thouësny (eds.). *CALL and professionalisation: short papers from EUROCALL 2021*, 1-5.
- Bédi, Branislav, Matt Butterweck, Cathy Chua, Johanna Gerlach, Birgitta Björg Guðmarsdóttir, Hanieh Habibi, Bjartur Örn Jónsson, Manny Rayner, Sigurður Vigfússon (2020). "LARA: An extensible open source platform for learning languages by reading". In: Karen-Margrete Frederiksen, Sanne Larsen, Linda Bradley, Sylvie Thouësny (eds.). *CALL for widening participation: short papers from EUROCALL 2020*, 27-35.
- Bédi, Branislav, Haraldur Bernharðsson, Cathy Chua, Birgitta Björg Guðmarsdóttir, Hanieh Habibi, Manny Rayner (2020). "Constructing an interactive Old Norse text with LARA". In: Karen-Margrete Frederiksen, Sanne Larsen, Linda Bradley, Sylvie Thouësny (eds.). *CALL for widening participation: short papers from EUROCALL 2020*, 20-26.
- Butterweck, Matt, Cathy Chua, Hanieh Habibi, Manny Rayner, Ghil'ad Zuckermann (2019). "Easy construction of multimedia online language textbooks and linguistics papers with LARA". In: *Proceedings of ICERI 2019*, 7302-7310.
- Lyding, Verena, Lionel Nicolas, Branislav Bédi, Karen Fort (2018). "Introducing the European NETwork for Combining Language Learning and Crowdsourcing Techniques (enetCollect)". In: Peppi Taalas, Juha Jalkanen, Linda Bradley, Sylvie Thouësny (eds.). *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018*, 176-181.

- Masuda, Kyoko (2018). *Cognitive Linguistics and Japanese Pedagogy: A Usage-based Approach to Language Learning and Instruction*. De Gruyter.
- Rayner, Manny, Matt Butterweck, Hanieh Habibi, Cathy Chua (2021). Constructing LARA Content. Online documentation. <https://www.issco.unige.ch/en/research/projects/callector/LARADoc/build/html/index.html>
- Schmid, Helmut (1994) “Probabilistic Part-of-Speech Tagging Using Decision Trees”. In: *Proceedings of the International Conference on New Methods in Language Processing*, 154-162.
- Steingrímsson, Steinþór, Örvar Káráson, Hrafn Loftsson (2019). “Augmenting a BiLSTM tagger with a morphological lexicon and a lexical category identification step”. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2019*, 1161-1168.

Elham Akhlaghi
ORCID iD: 0000-0002-4635-7780
Ferdowsi University of Mashhad
Azadi Square, Mashhad
Rhazavi Khorasan Province
Iran
elhamakhlaghi80@gmail.com

Anna Bączkowska
ORCID iD: 0000-0002-0147-2718
University of Gdańsk
Instytut Anglistyki i Amerykanistyki
ul. Wita Stwosza 51
80-308 Gdańsk
Poland
anna.baczkowska@ug.edu.pl

Branislav Bédi
ORCID iD: 0000-0001-7637-8737
The Árni Magnússon Institute
for Icelandic Studies

Skrifstofa Árnagarði við Suðurgötu
102 Reykjavík
Iceland
branislav.bedi@arnastofnun.is

Hakeem Beedar
Independent scholar
Adelaide
Australia
hbeedar@hotmail.com.au

Cathy Chua
ORCID iD: 0000-0002-5462-2441
Independent scholar
Adelaide
Australia
cathyc@pioneerbooks.com.au

Catia Cucchiarini
ORCID iD: 0000-0001-5908-0824
Centre for Language and Speech Tech-
nology (CLST)
Radboud University Nijmegen and
Union for the Dutch Language
Paleisstraat 9
2514 JA Den Haag
The Netherlands
c.cucchiarini@let.ru.nl

Hanieh Habibi
ORCID iD: 0000-0001-7602-2933
FTI/TIM
University of Geneva
40, Boulevard du Pont-d'Arve
1211 Geneve 4
Switzerland
hanieh.habibi@unige.ch

Ivana Horváthová
ORCID iD: 0000-0002-8388-7692
Constantine the Philosopher University
Department of English and
American Studies
Štefánikova tr. 67
Faculty of Arts, CPU
949 01 Nitra
Slovakia
ihorvathova@ukf.sk

Christèle Maizonniaux
ORCID iD: 0000-0003-2844-1820
Flinders University
College of Humanities, Arts and
Social Sciences
SSN3105, GPO Box 2100
Adelaide 5001
Australia
christele.maizonniaux@flinders.edu.au

Neasa Ní Chiaráin
ORCID iD: 0000-0002-4669-5667
Trinity College Dublin
Room 4074A Arts Building
College Green
Dublin
Ireland
neasa.nichiarain@tcd.ie

Nathalie Paterson
University of Adelaide
Faculty of Arts, Business, Law and
Economics (ABLE)
SA 5005 Adelaide
Australia
npa10716@bigpond.net.au

Chadi Raheb
ORCID iD: 0000-0002-5731-0443
University of Guilan

1841, Khalij Fars Highway
Rasht, Gullian
Iran
chadi.raheb@gmail.com

Manny Rayner
ORCID iD: 0000-0002-6002-2434
FTI/TIM
University of Geneva
40, Boulevard du Pont-d'Arve
1211 Geneve 4
Switzerland
emmanuel.rayner@unige.ch

Chunlin Yao
ORCID iD: 0000-0001-6006-5031
Tianjin Chengjian University,
No. 26 Jinjing Road, Xiqing
Tianjin
China
yao_chunlin@126.com

Teaching sustainability communication in higher education: Applying case study method based on social media

DALIA GULBINSKIENĖ
GIEDRĖ VALŪNAITĖ OLEŠKEVIČIENĖ

*Received 27.08.2021,
received in revised form 23.11.2022,
accepted 24.12.2022.*

Abstract

The article aims to introduce the role of foreign language teaching while using social media in fostering sustainability communication (SC) in higher education (HE) institutions. Foreign language learning enables students to analyse actual topics about sustainable development on social media, to discuss problems, to share personal emotions and experiences. Moreover, ESP curriculum aims to build students' basic language communicative skills with the focus on sustainability communication when students recall gained information. The article overviews the importance of the SC development on the theoretical level as well as introduces the challenges of SC development relying on social media texts in foreign language classes on the empirical level. One of the most innovative methods of learning/teaching a foreign language is a case study method. It integrates such skills as one's ability to analyse, think critically, search for solutions and make decisions. The article deals with concepts of professional English development and case analysis based on social media. It presents the findings of the research into students' attitude towards case study method based on social media texts for developing their professional English commu-

nication. The research highlights the problems and peculiarities of applying the method. It has revealed that students have a positive attitude towards assignments involving case study method based on social media texts. They believe that such assignments help develop one's communication skills of the professional English language.

Keywords

sustainability communication (SC), social media, English for Specific Purposes (ESP); higher education (HE), case study, communication skills

Nauczanie komunikacji dotyczącej zrównoważonego rozwoju w szkolnictwie wyższym: zastosowanie podejścia studium przypadku opartego na sieciach społecznościowych

Abstrakt

Artykuł ma na celu przybliżenie roli nauczania języków obcych z wykorzystaniem portali społecznościowych w promowaniu zrównoważonej komunikacji (SK) w szkołach wyższych. Nauka języków obcych daje możliwość analizowania aktualnych tematów zrównoważonego rozwoju, omawiania problemów oraz dzielenia się osobistymi emocjami i doświadczeniami w sieciach społecznościowych. Ponadto, program nauczania ESP ma na celu rozwijanie podstawowych umiejętności komunikacyjnych uczniów, z naciskiem na trwałą komunikację, w której uczniowie zapamiętują informacje, których się nauczyli. Artykuł dokonuje przeglądu znaczenia rozwoju SC na poziomie teoretycznym, a także przedstawia wyzwania związane z rozwojem SC w oparciu o teksty z mediów społecznościowych na lekcjach języków obcych na poziomie empirycznym. Jednym z najbardziej innowacyjnych sposobów uczenia się/nauczania języka obcego jest metoda studium przypadku. Integruje takie umiejętności, jak umiejętność analizowania, krytycznego myślenia, znajdowania rozwiązań i podejmowania decyzji. W artykule przeanalizowano koncepcje profesjonalnego rozwoju języka angiels-

kiego i analizę przypadku w oparciu o sieci społecznościowe. Przedstawiono również wyniki badania postaw uczniów wobec opartego na tekście studium przypadku w mediach społecznościowych w celu rozwijania profesjonalnej komunikacji w języku angielskim. W pracy zwrócono uwagę na problemy i specyfikę zastosowania metody. Okazało się, że studenci mają pozytywne nastawienie do zadań, które wykorzystują podejście studium przypadku oparte na tekstach z mediów społecznościowych. Poza tym, studenci wierzą, że takie zadania pomagają rozwijać umiejętności profesjonalnej komunikacji w języku angielskim.

Słowa kluczowe

komunikacja zrównoważonego rozwoju (SC), media społecznościowe, angielski do określonych celów (ESP); wykształcenie wyższe (HE), studia przypadków, umiejętności komunikacyjne

1. Introduction

Higher education institutions play a significant role in the attainment and demonstration of sustainability communication in which proficiency in a foreign language is an essential prerequisite for acquiring desired results for sustainability communication competence development. In today's globalised world there is an increasing need for students to improve foreign language competences at university. There has been an ongoing discussion on how to help students be more adaptable to the new environment and more focused on sustainability communication (SC). According to Godemann (2011), Tilbury (2011), Iwaniec et al. (2014), the four-dimensional (economical, environmental, social and institutional) the concept of sustainable development has been determined. Furthermore, the fourth SC dimension, which is related to the complex challenges of contemporary society, has modified the SC structure including institutional dimension and emphasizing an autonomous activity, the use of interactive foreign language through participation in

dialogues for the present and future of global society. These constructs are used in different environments and contexts as they are assumed to be significant elements of communication. Case study method based on social media is a major tool in teaching sustainability communication in higher education.

Case study method is based on descriptions of real decisions made in the real world, which were introduced to learners as cases. Learning/teaching based on this method focuses on cooperative learning principles grounded in the mechanism of self-regulation as well as using cases in conceptual, analytical, and presentation learning dimensions. It develops the main eight transferred abilities, namely analytical thinking, decision-making, application, oral communication, time-planning, interpersonal/social, creative, written communication. Currently case study method, already acknowledged as the alternative to traditional forms of learning and teaching, is gaining more popularity. Case study method is regarded as the method which develops one's communication competence. It is a method based on active participation, cooperation, and democratic discussion within a group. Discussion, if applied appropriately, encourages acquisition of knowledge, abilities and approaches. There is no unanimous method of case study. There are plenty of them to be applied in different situations for dealing with various problems.

Research object is teaching sustainability communication in higher education by using social media. Research aim is to identify the importance of case study method when teaching sustainability communication skills by using social media in higher education. Research objectives are as follows: a) investigate the phenomenon of sustainability communication; b) highlight the impact of case study method based on social media and c) to find out students' opinion about the use of case study method in training sustainability communication by using social media. Research methods are the analysis of scientific literature was conducted in order to theoretically base the concepts of case study based on social media; written survey, which helped identify what impact case study method makes on the development

of students' sustainability communication skills in higher education.

2. Theoretical background

Sustainability communication is a wide and complex interdisciplinary phenomenon and covers many social, economic, environmental, institutional, emotional and spiritual spheres of human life. Moreover, the internal structure of sustainability communication (knowledge, talents, abilities, skills, moral values, attitudes, behavioural intentions and their interrelations) is profound as well as external elements – institutional, social, economic and environmental dimensions – emphasizing their role which they play in designing the needs to be acquired by students in higher education institutions. However, researchers have not established the finalized SC definition or unifying theory and concept. For this reason, interpretations are different, based on ideological and educational perspectives involving aims related to the application of the concept both in scientific and policy making discussions. Godemann (2011) defined sustainability communication as a complex process of information and knowledge exchange between sender and receiver. Its structure comprises language, knowledge, abilities, skills, moral values, potential, and talents, all of which allow learners to achieve mutual understanding in communication and undertake various activities or tasks (which is referred to as competency) related to challenges of sustainable development and education. What is more, the notion of sustainability communication might be referred to the necessity and desire to fulfil the needs of leadership, professional career, social status, and/or other personal achievements. Furthermore, it is also significant to figure out if learners, who have some targets, are able to establish a possession of certain level of competences to overcome obstacles in order to reach their goals. Therefore, it is important not to limit the development of competences to their cognitive elements and learners should be informed about internal elements of a desired competence, too. This conceptualization is essential for

sustainability communication as it is based on holistic and humanistic philosophy combining all necessary elements such as moral attitudes, internal and external motivation, learning environment and mentors' support that are significant for an effective performance. Therefore, pedagogues have a very important mission in empowering students to contribute to a more sustainable world while equipping them with the ideas of sustainability communication as a forceful and useful tool. The sustainable development paradigm ensures a concrete context of university mission and vision signifying a new stage of study quality achievement. Moreover, based on developments by European Commission (2017) researchers of the institutionalism (Filho et al. 2013) and social constructivism theories, Tilbury (2011) analysed activities of institutions, their structure, formal rules and informal rituals as well as relationships between higher education institutions and their importance. A university is an organizational unit which foresees and ensures university sustainable development and sustainability communication at all institutional levels. The university carries out social, economic, environmental politics: uses energy, transport and performs other activities related to infrastructure. In order to improve the learning environment of students, the university infrastructure is constantly renewed.

Highlighting the place of English for Specific Purposes (ESP) in teaching sustainability communication, Sterling (2004) points out that 'it requires an understanding of the role that language and culture play in the construction of environmental, social, economic, institutional, cultural and religious systems, the impact of these systems which support life'. Therefore, it could be stated that ESP might ensure students to become more motivated to develop their sustainability communication through English learning activities. The problem is how to foster the development of sustainability communication through the ESP curriculum within the context of tertiary level. In the system of higher education of Lithuania, the ESP curriculum aims to build students' language skills, to improve their knowledge about sustainability communication and to enable them to apply sustain-

nable development ideas with respect for the environment, in which cultural reproductions have major implications not just for the content of the curriculum, but for forms of pedagogic interactions. For this reason, it is challenging for pedagogues to believe that ESP will improve students' sustainability communication.

Today university students are supposed to gain skills of independent work, be able to apply their skills and knowledge in newly encountered situations, use social networks and social media, and up to date technologies. European Commission document states that linguistic competences are a part of general abilities, which ensure opportunities of each citizen's employment, education and personal improvement. These abilities should be regularly renewed and replenished. When teaching foreign languages, we develop learners' social and cognitive performance and creativity. When communicating with people, students are supposed to take into account persuasiveness, culture and style of their language and being able to use the newest sources of social media. Inefficient communication can cause the problems such as lack of innovations, poor performance, failure in the case of a problem, and threats when dealing with it. Employers, who have faced communication problems, try to improve their sustainability communication skills, correct their own mistakes and eliminate drawbacks. For this reason, not only, existing methodologies are used but also new ones are designed.

Aiming to equip students with the possibility to develop their sustainability communication competence is a big challenge of higher educational institutions. When presenting the study programmes, it is important to introduce students with the concept of sustainability communication, its elements, teaching methods and learning strategies that could be used to provide effective sustainability communication competence development through ESP learning processes. According to Godemann (2011), Flowerdew (2013) and Wiek, et al. (2014), it could be claimed that the content of sustainable development is connected with communicative (native and foreign language skills),

cross-cultural (foreign language skills and knowledge about different cultures) personal (ability to manage internal and external elements), methodological – instrumental (knowledge and skills of a particular profession), social (demonstrating ability of environment protection), transformative (demonstrating ability of positive change) as well as core competences (IT, mother tongue, mathematics and etc.). Similarly, to sustainable development competence and its components, sustainability communication is strongly influenced by social and mass media and therefore has a variety of characteristics.

The characteristics of sustainability communication encompass reflexivity in relation with various problems; successful ability to deal with difficult situations, medialisation as an effort to counter the tendency to normalization in sustainability discourse by matching it with the forms of mass media. Tendencies to normalization with the possibility that the more sustainability becomes a topic and norm, the less stress and pressure there is to reach mutual understanding. Establishment of sustainability as an intrinsic social value and the related issue of creating agreement, the possibility of various interpretations of sustainability have significant influence and should be taken into considerations when curriculum development processes are made and ESP teaching processes are designed. Moreover, social networks, made up of individual and group actors with their “bridges”, are increasing. The examples in the mediated forms of communication and social interaction processes are part of these ties with their own specific online practices, rules, network relationships and technical possibilities. Therefore, handling this complexity of interdisciplinarity plays a crucial role in the discussion about sustainability. Moreover, communication about sustainable development is sustainability communication about sustainable development knowledge and experiences. However, simply emphasizing the importance of the concept of sustainability is not enough to raise students’ awareness about sustainable development. Knowledge needs a practical value, for this reason systemic knowledge must be acquired, i.e., the knowledge of functions, processes and interrelationships. Sust-

ainable development knowledge needs to be assimilated with values, ethical orientations towards the links between humans and nature, with direct experiences that involve emotionality and meaningfulness. ESP teaching has the medium and a long-term goal to assist students not only to improve their foreign language skills, but also to acquire basic knowledge about sustainability to actively communicate its ideas and link them with actions. Teaching ESP in higher education aims at developing and enhancing communication about sustainability that allow students to sharpen their awareness in both private and working life of what is ecologically responsible, economically feasible, socially acceptable as well as enabling them to discuss their ideas in classroom, using and analysing ideas expressed on social media. SC is classified as a “soft” tool, and is one of a number of popular instruments, therefore, involving students in the solution of their own problems opens up opportunities for them making corresponding changes in their behaviour for influencing the present and future life.

Many teachers have shared their experience of using case study method when learning/teaching foreign languages. This method is frequently applied in higher education- video examples of simulation cases are provided and discussed, then real cases are selected, i.e., cases which have free access on social media or simulation cases included in specialized publications. Students are divided into groups and analyse the material of the selected case in detail. It is proceeded by acting of the selected case. Case study method is crucial when teaching sustainability communication in higher education.

The use of case study method when teaching/learning ESP positively affects development of active and passive linguistic skills as it is an integrated method. It develops other skills and competences as well, for instance, problem solution, presentation-making, team work, analytical skills. By providing students with a possibility to make decisions, their knowledge in the area under discussion is enriched. They are encouraged to understand the complexity of the situation.

What is more, the use of authentic material of social media enhances teaching/learning experience (Kop 2010; Oleskeviciene 2020) as social media keeps the students engaged and encourages their participation as students not only use the material but also express themselves on social media by publishing their work and discussions. It also encourages their collaboration as it enhances student interaction and information sharing. Besides, social media enables the teachers themselves to share their materials, worksheets or websites.

3. Research methodology

The research was conducted at the Faculties of Environmental Engineering and Creative Industries at Vilnius Gediminas Technical University. 40 first year students participated in the research. They were coded as students S1-S40 in the research findings below. Respondents were allowed to choose a colleague to work in pairs. The pairs selected the topics for case study on social media. Firstly, they had to analyse the case of sustainability communication, think of solutions and provide recommendations. Two weeks were given for the assignment. Students had to present their assignment in public: introduce their group friends with the case analysed, provide options for solutions and publish their work on social media and collect the feedback. Respondents could individually select the forms of visual aids. The teacher encouraged a creative approach towards the assignment and quest for brave solutions, emphasizing that there are no incorrect answers. The main attention was paid to a foreign language. During the course students learn English terms, using all types of linguistic activity, namely reading, writing, speaking and listening. Research participants were warned that primarily they were evaluated for argument-based solutions, well-made presentation in English using sustainability terminology as well as application of public speech rules and preparation of visual aids and materials for publishing on social media.

The following research questions were formulated according to the theoretical assumptions discussed above and personal insight:

- How are language skills developed when learning via the case study method using social media?
- How does case study and using social media promote social interaction?
- What new skills do students acquire through the case study method using social media?

In order to find out the answers to the research questions, a questionnaire was developed, which included closed-ended and open-ended questions to obtain quantitative and qualitative data. The questionnaire was provided to students who have already had one semester of experience of studying English for Specific Purposes at the university and the case study method was applied in these groups. The questionnaire was submitted to 40 respondents and they were informed that it aimed to summarize their authentic experiences in the language learning process and understand what they think about case study method using social media, as well as use their insights to improve the learning process in the future.

4. Research findings and discussion

According to the results of the student survey, as many as 100 % of the respondents answered that case study method using social media and participation in group discussions create preconditions for a favourable learning environment. According to one of the respondents: 'Case study method using social media ensures a good atmosphere, you can learn from others', another comment: 'Case study method using social media creates a favourable atmosphere for learning, because in a group you learn to collaborate with others, to work in a team'.

Moreover, 95 % of the respondents answered the question whether case study method using social media improves the

relationship between group members and the relationship with the teacher in the affirmative way, stating in the comments that students communicate more with each other and with the teacher, and only 5 % thought that this is not always the case because sometimes disputes can arise, especially when opinions differ widely. Despite possible differences in opinions, 95 % of the respondents said that case study method using social media and group work help learners to get to know and understand other people better. One of the comments stated that ‘while working in a group in chat rooms, you do tasks with different people, try to communicate, listen, and understand others.

Regarding the development of language skills through case study method using social media, 89 % of the respondents indicated that they developed language skills more efficiently, 74 % also mentioned listening skills and only 16 % thought that they also developed reading skills, while only 11 % mentioned writing skills (Figure 1). Respondents also mentioned other skills such as the ability to discuss, the ability to negotiate, and tolerance.

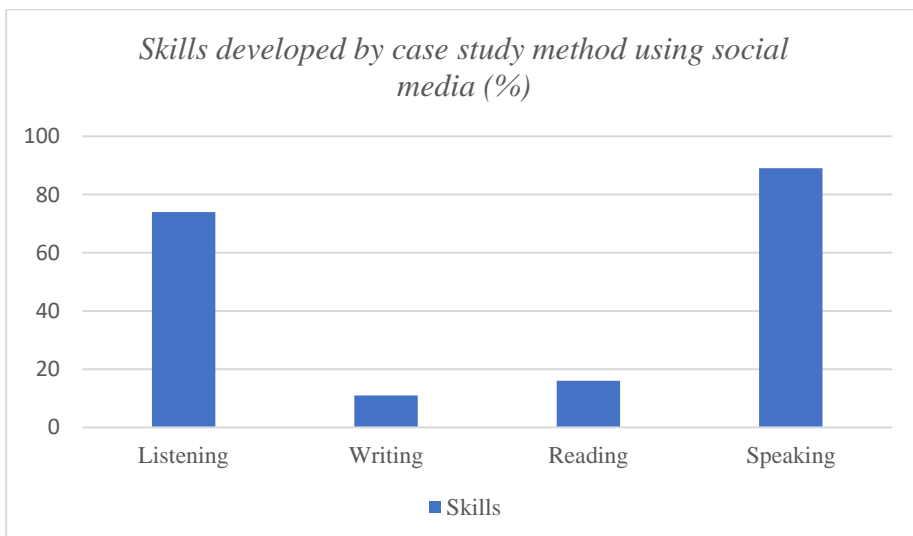


Figure 1

Skills developed by case study method using social media

For comparison, we asked a question about the skills that students develop when performing tasks individually. The majority, i.e., 79 %, of the respondents indicated reading, 30 % of the respondents mentioned listening, 23 % of the respondents mentioned writing and only 13 % mentioned speaking (Figure 2). In addition, none of the respondents indicated any additional skills.

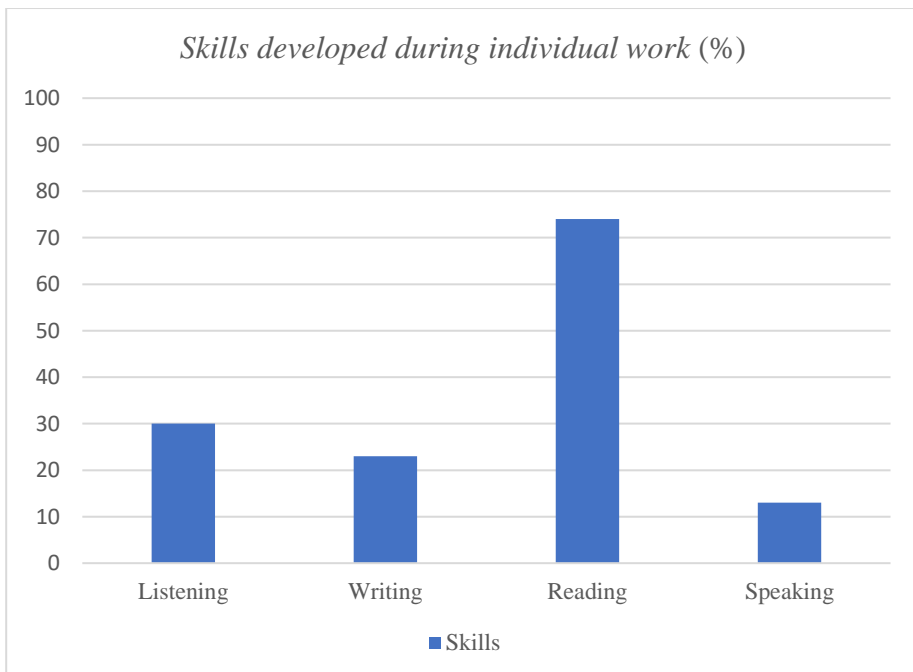


Figure 2

Skills developed during individual work

Comparison of the results of the responses to these two questions shows that case study method using social media significantly develops speaking skills, thus this method and group work approach should be used more often. This means teachers should practice arranging student collaborative work in chat rooms.

Students were also asked about social skills that they developed through the case study method using social media. 100 % of the respondents indicated that they learnt how to express

their opinion. 89 % of the respondents said they were able to agree or disagree with a colleague. 79 % of the respondents indicated the ability to listen to a colleague, 25 % of the respondents learnt how to praise and encourage, and 17 % of the respondents indicated the ability to intervene into the discussion (Figure 3).

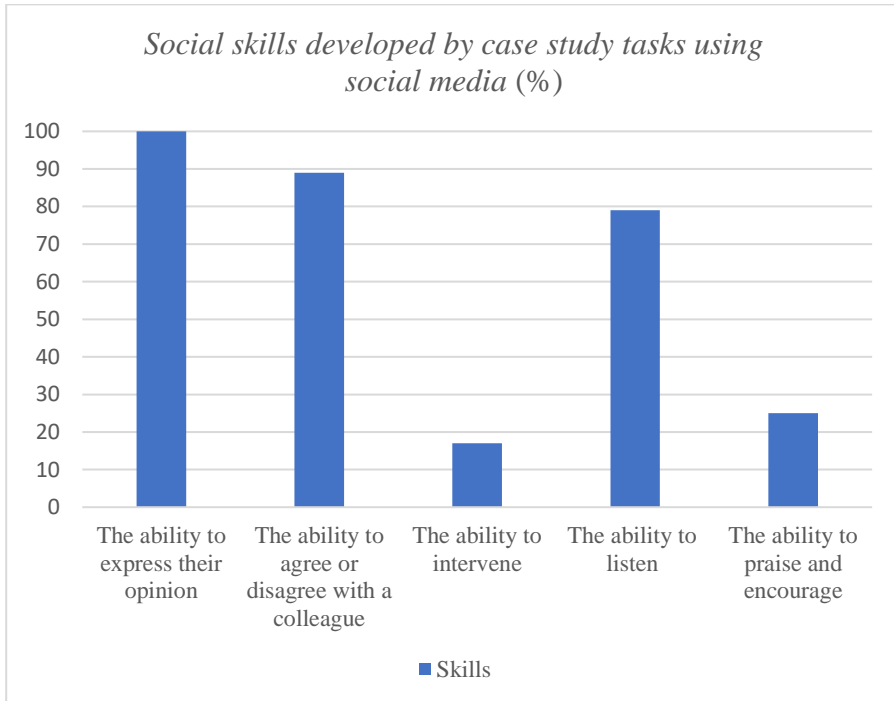


Figure 3

Social skills developed by case study tasks

When asked what they like about the case study method using social media, the respondents mentioned such factors as communication, listening, expressing their opinion, new ideas expressed by the other students, help, and cognition of colleagues as personalities.

When asked what disadvantages the case study method using social media might have, the majority of the respondents, i.e., 94 % said that they liked learning in this way very much and only a few indicated that there could be disputes, inter-

ruptions and lack of attention. Finally, when asked what activities they would like during the seminars, the majority of the respondents, i.e., 89 %, answered that they would like to learn in this way.

To sum up the results of the survey, we can state that they demonstrate that the case study method significantly improves the development of speaking skills, as well as learners' social skills such as being able to express one's opinion, agree or disagree with colleagues, intervene, listen, praise or encourage; it helps to create a learner-friendly atmosphere, improves interpersonal relationships in the group and is attractive to learners, as most respondents indicated that they would like to learn in a this way.

The research findings section below presents the thematic structure of the phenomenon of applying case study method based on social media in university studies. The process of clustering the statements into subthemes and themes is presented in the research findings together with the analysis of the phenomenon of case study method using social media. While analysing the empirical interview material, two main themes were identified: the benefits of case study method integrated with social media and the challenges of case study method integrated with social media.

The theme of the benefits of case study method integrated with social media splits into subthemes of using social networks and information availability, peer interaction and collaboration, improving English and presentation skills. This theme reveals the skills which are perceived by the research participants as substantially improved during their university studies and applicable in the workplaces. The research participants stress the improved ability to use social media, to interact and collaborate when learning, to solve problems and improve English skills acquired during case study activities. The ability to use new technologies, which is closely related to information processing skills, is also mentioned by the research participants. Overall, the subthemes are related to one of the general domains of the sustainability communication framework which is the media

and information literacy. The following subthemes of the capacity to organize one's ideas, decision making and being able to solve problems are related to the domain of interpersonal skills which are also of the utmost importance in work life and are perceived as substantially improved when applying case study method (see Table 1).

Table 1

The respondents' feedback on the benefits of case study method

Theme	Sub-theme 1	Sub-theme 2	Meaning Unit
I. The benefits of case study method integrated with social media	Using social media	Engaging activities	<i>S14, S20, S23 "<...>the task was engaging as I had to use social networks"</i>
		Information availability	<i>S6, S10, S11 "I found out a lot about sustainability communication on social media<...>" S40 "<...>the assignment was useful because I learned about successful companies"</i>
	Interaction	Learning problem solving	<i>S1, S5 "it encouraged to look for solutions <...>".</i>
		Learning collaboration	<i>S13, S15 "<...>it was much more interesting to work together with my group mate than individually"</i>
		Improving English skills	<i>S17, S12, S39 "I used professional terminology<...>"</i>

	Skills	Improv- ing presenta- tion skills	S2, S19, S37 " <...> I improved my skills of making presentations and publishing material"
--	--------	---	--

It could be seen in Table 1 that the research participants express their willingness to apply case study method using social media during our ESP classes. In the first subtheme research participants stress the need and importance of using social media when learning about sustainability communication. *"The task was engaging as I had to use social networks"* S14, S20, S23.

They also acknowledge the importance of improving English skills which enable sustainability communication. The importance of interaction stands out in the research as the research participants stress the benefits of cooperation skills. *"It was much more interesting to work together with my groupmate than individually"* S13, S15. Research participants also recognize the importance of critical thinking which is of crucial importance in the developing knowledge society, thus the respondents' feedback about case study method is very positive *"It encouraged to look for solutions <...>"* S1, S5.

The theme of the challenges of case study method integrated with social media reveals the skills, which are perceived as necessary in work life, and the need for improvement.

It could be seen that the research participants express their willingness for more improvement of certain skills. First, in the subtheme of the need for critical thinking research participants stress the need and importance of analysing and problem-solving skills. They also acknowledge the importance of foreign languages which enable cross-cultural communication. The importance of subject knowledge stands out in the research. Research participants also recognize the importance of the ability to cooperate and collaborate, the skills which are of crucial

importance in the developing knowledge society. Learning to learn and time management are perceived as useful both in study and work environments and the need for their development is expressed (see Table 2).

Table 2

The respondents' feedback on the challenges of case study method

Theme	Sub-theme 1	Sub-theme 2	Meaning Unit
II. The challenges of case study method integrated with social media	Need for critical thinking	Need for analysing skills	<i>S18, S6 "<...>it was difficult to conduct analysis, I needed some help"</i>
		Need for problem solving skills	<i>S10, S16"<...>it was hard to think of solutions and provide recommendations"</i>
	Need for subject knowledge	Need of English skills	<i>S8, S11, S38 "it was a challenge to understand the text in English <...>".</i>
		Need for professional knowledge	<i>S9, S21, S32"<...>I do not have enough professional knowledge on sustainability and engineering, I need to improve it "</i>

	Peer communication	Need for discussion time	<i>S17, S12, S25 " <...> we needed more time for discussing the assignment"</i>
		Need for better collaboration	<i>S3, S22 " <...> I am not satisfied with my partner's activity, I expected a better cooperation"</i>

As can be seen in Table 2, research participants also identify the need to shape focus more on professional English which could be applied in work life in practice *"I do not have enough professional knowledge"* S9, S21, *"It was hard to think of solutions and provide recommendations"* S10, S16. Despite some challenges of applying case study method, the identified needs by the research participants highly resonate with the needs to develop sustainability communication in university studies discussed by the researchers working in the field (Godemann, 2011, Tilbury 2011, Iwaniec et al. 2014).

5. Conclusions

Sustainability communication is a complex, diverse and mult disciplinary concept, therefore sustainability communication challenges could be analysed emphasizing the involvement of all language pedagogues for the implementation of sustainability communication at university. It is crucial to involve language teachers in solutions to sustainability communication problems, plan, coordinate and perform activities for the development of sustainability communication. English language learning makes a positive impact on students' knowledge and skills about sustainability communication. Sustainability communication education is not the phenomenon which is taken for granted, but on the contrary, has to be developed by pedagogues

and supported by students at higher education institutions. Sustainability communication expands the spectrum of English language teaching and promotes students' opportunities of foreign language learning at all stages at higher education. The practical significance of the research is that it addresses students' needs for the education of sustainability communication and sustainable development that affects students' efficacy of acquisition of sustainability communication knowledge and skills. Embedding sustainability communication content and innovative methods of teaching into ESP classes, pedagogues are encouraged to update their teaching strategies with innovative approach to discuss sustainability communication. The possibility to apply this novelty for English language learning is estimated on empirical study results and such practice is proved to be successful for the development of students' linguistic skills and knowledge about sustainability communication. Moreover, the acquisition of sustainability communication through English language learning is the privileged experience gained while studying at universities, and case study method based on social media is the best tool in teaching sustainability communication in higher education. Theoretical and practical provisions of the research revealed that despite some challenges, case study based on social media is a relevant method when teaching sustainability communication at university level. Case study method helps to develop professional English competence, requires analytical skills, provides numerous solutions and teaches students think positively. Furthermore, it improves knowledge in the area of professional training and helps perceive complexity of the situation. Students learn to describe the existing situation, identify problems, provide arguments and make decisions. They are also able to discuss advantages and disadvantages of selected decisions and draw conclusions. The research has shown that case study is one of the most efficient methods, which helps establish skills of professional English in an easier and more rapid way.

Our research reveals that students' social skills as well as their English language skills are substantially improved by app-

lying case study method based on social media. They especially stress such skills as the ability to express their opinion, the ability to agree or disagree with a colleague, the ability to speak English, the ability to use new technologies, which proves that sustainability communication is of key importance in our knowledge society. Overall, these skills are related to one of the general domains of the sustainability communication framework which is the media and information literacy. The research participants also identify the capacity to use social media, to collaborate with peers, to solve problems, organize one's ideas, make decisions which are important in work life. The research participants express their perceived needs for developing certain skills which are necessary in professional work life. The research participants stress the need and importance of critical thinking and analysing skills, subject knowledge, peer communication, English skills and professional knowledge which enable sustainability communication, which resonates with the ideas expressed by the research carried out in the domain (Godemann (2011), Flowerdew (2013) and Wiek, et al. (2014)).

References

- Balčiūnaitienė, Asta. (2019). "Development of sustainability communication in EFL classroom at higher education". *Public Security and Public Order* 2019 22.
- Beckisheva, Tatyana. G. (2014). "Case Study as an Active Method of Teaching Business English". *Procedia. Social and Behavioral Sciences* 166/7: 292-295.
- European Commission (2017). "Recent Developments in European Higher Education Systems. European Commission staff working documents (SEC, 1063 final)". Retrieved 21/04/2019 from <http://ec.europa.eu/education/pub/pdf/higher/modernisation.pdf>
- Filho, Leal Walter, Sandra Caeiro, Charbel Jabbour, Ulisses M. Azeiteir (2013). *Sustainability Assessment Tools in Higher Education Institutions*. New York: Springer International Publishing.
- Filho, Leal Walter (2015). "Education for sustainable development in higher education". In: Walter Leal Filho (ed.). *Transformative*

- Approaches to Sustainable Development at Universities*. New York: Springer International Publishing, 3-10.
- First European Survey on Language Competences (2012). European Commission. Education and Training
- Flowerdew, John. (2013). *Discourse in English Language Education*. New York: Routledge.
- Iwaniec, David M., Daniel L. Childers, Kurt VanLehn, Arnim Wiek (2014). "Studying, teaching and applying sustainability visions using systems modelling". *Sustainability* 6/7: 4452-4469.
- Kop, Rita (2010) *Networked Connectivity and Adult Learning: Social Media, the Knowledgeable Other and Distance Education*. Doctoral thesis. Swansea University.
- Oleskeviciene, Giedre Valunaite, Jolita Sliogeriene (2020). *Social media use in university studies*. New York: Springer.
- Stepanovienė, Aušra (2017). "Applying case study method while developing communicative competence of professional English". *Public Security and Public Order* 18: 230-237.
- Sterling, Stephen (2004). "An analysis of the development of sustainability education internationally: Evolution, interpretation and transformative potential". In: John Blewitt, Cedric Cullingford (eds.). *The Sustainability Curriculum: A Challenge for Higher Education*. London: Earthscan, 43-62.
- Tilbury, Daniella (2011). "Are we learning to change? Mapping global progress in education for sustainable development in the lead up to "Rio Plus 20"". *Global Environmental Research. Education for Sustainable Development: Promises and Challenges* 14/2: 101-107.

Dalia Gulbinskienė
ORCID iD: 0000-0002-1284-795X
Department of Foreign Languages
VilniusTech
pl. Saulėtekio 11
LT-10223 Vilnius
Lithuania
dalia.gulbinskiene@vilniustech.lt

Giedrė Valūnaitė Oleškevičienė
ORCID iD: 0000-0001-5688-2469
Department of Foreign Languages
Mykolas Romeris University
pl. Ateities 20
LT-08303 Vilnius
Lithuania
gvalunaite@mruni.eu

Beyond Philology No. 19/1, 2022
ISSN 1732-1220, eISSN 2451-1498

Information for Contributors

Beyond Philology is an international journal of linguistics, literary studies and language acquisition. The journal publishes articles, reviews, reports and interviews. The language of the journal is English.

The preferred length of papers is 3500 words, but shorter and longer manuscripts are also considered for publication. Manuscripts should be accompanied by abstracts of about 150 words and keywords.

Each contributor should send a note stating specifically that his/her manuscript is only being considered for publication in *Beyond Philology*.

Manuscripts (papers, book reviews, reports, interviews) and books submitted for review should be sent to:

beyond.philology@ug.edu.pl

For more information see:

https://fil.ug.edu.pl/wydzial_filologiczny/instituty/institut_anglistyki_i_amerykanistyki/czasopismo_naukowe_beyond_philology
<https://czasopisma.bg.ug.edu.pl/index.php/beyond>

All correspondence concerning subscriptions should be sent to:

Wydawnictwo Uniwersytetu Gdańskiego
ul Armii Krajowej 119/121
80-824 Sopot, Poland
Phone/fax: (48) 58 523 11 37
<http://www.wyd.univ.gda.pl>
Email: wyd@ug.gda.pl