

ANGLICA

An International Journal of English Studies

25/2 2016

EDITOR

prof. dr hab. Grażyna Bystydzieńska [g.bystydzienka@uw.edu.pl]

ASSOCIATE EDITORS

dr hab. Marzena Sokołowska-Paryż [m.a.sokolowska-paryz@uw.edu.pl]

dr Anna Wojtyś [a.wojtys@uw.edu.pl]

ASSISTANT EDITORS

dr Katarzyna Kociołek [kkociolek@uw.edu.pl]

dr Magdalena Kizeweter [m.kizeweter@uw.edu.pl]

ADVISORY BOARD

Michael Bilynsky, University of Lviv
Andrzej Bogusławski, University of Warsaw
Mirosława Buchholtz, Nicolaus Copernicus University, Toruń
Xavier Dekeyser, University of Antwerp / KU Leuven
Bernhard Diensberg, University of Bonn
Edwin Duncan, Towson University, Towson, MD
Jacek Fabiszak, Adam Mickiewicz University, Poznań
Jacek Fisiak, Adam Mickiewicz University, Poznań
Elżbieta Foeller-Pituch, Northwestern University, Evanston-Chicago
Piotr Gąsiorowski, Adam Mickiewicz University, Poznań
Keith Hanley, Lancaster University
Christopher Knight, University of Montana, Missoula, MT
Marcin Krygier, Adam Mickiewicz University, Poznań
Krystyna Kujawińska-Courtney, University of Łódź
Zbigniew Mazur, Maria Curie-Skłodowska University, Lublin
Rafał Molenccki, University of Silesia, Sosnowiec
John G. Newman, University of Texas at Brownsville
Michał Jan Rozbicki, St. Louis University
Jerzy Rubach, University of Iowa, Iowa City
Piotr Ruskiewicz, Pedagogical University, Cracow
Hans Sauer, University of Munich
Merja Stenroos, University of Stavanger
Krystyna Stamirowska, Jagiellonian University, Cracow
Jeremy Tambling, University of Manchester
Peter de Voogd, University of Utrecht
Anna Walczuk, Jagiellonian University, Cracow
Jean Ward, University of Gdańsk
Jerzy Wehna, University of Warsaw

GUEST REVIEWERS

Magdalena Bator, University of Social Sciences, Warsaw
Magdalena Charzyńska-Wójcik, John Paul II Catholic University of Lublin
Joanna Esquibel, SWPS University
Anna Fornalezyk-Lipska, University of Warsaw
Barry Keane, University of Warsaw
Marta Kisielewska-Krysiuk, University of Warsaw
Jarosław Krajka, Maria Curie-Skłodowska University in Lublin
Jadwiga Linde-Usiekiewicz, University of Warsaw
Sylwester Łodej, Jan Kochanowski University, Kielce
Zbigniew Możejko, University of Warsaw
Mieczysław Nasiadka, University of Warsaw
Jerzy Nykiel, University of Bergen
Monika Opalińska, University of Warsaw
Agnieszka Piskorska, University of Warsaw
Dorota Rut-Kluz, University of Rzeszów
Marta Sylwanowicz, University of Social Sciences, Warsaw
Piotr Szymczak, University of Warsaw
Anna Warso, SWPS University
Iwona Witczak-Plisiecka, University of Łódź



UNIVERSITY
OF WARSAW

Interference Patterns of the Verb *Say* in the Narration of English-Polish Literary Translations: A Corpus-Based Study

Abstract

In this paper I present a corpus-based study of the interference of the verb *say* from English source texts into their Polish translations. I compare a sizeable corpus of native Polish fiction with corpora of English-Polish fiction translations and argue that the various forms of the verb *say* commonly used in the English narration influence the choices made in some (albeit not all) of the Polish translations. In order to exemplify my claims I use parallel corpora to show how two different professional translators working with an identical set of source texts were influenced by the narrative verb distribution.

1. Introduction

Baker (2001, 206) hypothesizes that the translation process involves stylistic and syntactic interference. Thus, a translator does not deal with the source text alone but also struggles to defy the influence of the source language. Before corpus tools became accessible the translation research was mainly focused on deeper layers of the text: by examining semantic and pragmatic equivalence researchers analyzed how well the content of source texts transitions to target texts. The research on syntax was qualitative, mostly focused on tangible items such as calques and borrowings.

Contemporary corpus tools make it possible to implement a much deeper analysis of texts. The data obtained from corpus research are qualitative, vast and provide insights not available to the naked eye. Instead of examining the texts piece by piece, a corpus tool processes them into a database – a corpus, which can then be browsed and examined (Crystal 2003, 91). Even though the corpus alone does not reveal anything specific about the text, its proper investigation exposes patterns and peculiarities present in the given text or set of texts. This, in turn, provides the researchers with very accurate statistical data. By no means do the corpus tools invalidate other approaches to analysing translations. They do, however, provide a parallel point of view, allowing the linguists to embrace the less visible aspects of the text.

Corpus methods are now widely used by translation theorists. Mauranen (2004, 72) uses research on multicompetence (Cook 2003, 18) as a justification for the source text-target text relationship appearing in literary texts translated into Finnish. Mauranen's corpus research focused on lexical distribution and the data indicated significant differences between native Finnish texts and texts translated into Finnish. Nilsson (2004, 5) discovers the overuse of the Swedish preposition *ov* 'of' in English-Swedish translations. The corpus data and follow-up qualitative research suggest that translators have a tendency to use *ov* in place of the English *of* even when more salient Swedish expressions are available. Fokin (2015, 6) uses keyness and TTR tests (see section 5. for definitions) to compare the distribution of lexical words and inflections in native texts with translations across a number of languages, discovering visible inconsistencies.

2. Concept of lexical equivalence and its applications

When investigating translators' decisions, one inevitably has to consider equivalence. This concept refers to the intricate relationship between various aspects of the source text and the target text (hereinafter abbreviated to ST and TT respectively) as well as the source language and the target language (SL and TL). Nord ([1997] 2014, 35) defines it as "a static, result-oriented concept describing a relationship of 'equal communicative value' between two texts" or smaller components of those texts. Since any two given languages will be inherently different, it is the translator's responsibility to take those differences into account and make appropriate choices (Pym 1992, 37). Baker (1992, 5) proposes to distinguish specific types of equivalence depending on the layer of the text they apply to (e.g. word-level or lexical equivalence, pragmatic equivalence, grammatical equivalence, etc.). These layers are co-dependent and inconsistencies on one layer may impact the others.

This paper focuses on lexical equivalence, which involves the choices with regard to the use and distribution of words. The translator's knowledge is thus tested extensively as they endeavour to adhere to TL norms: appropriateness of collocations, grammatical constraints and accuracy (Baker 1992, 48). Lexical equivalence is hence inherently connected to the linguistic phenomenon of anisomorphism, which assumes that perfect equivalence on the lexical level simply does not exist. Instead, seemingly equivalent words will always differ with regard to their distribution, salience, connotations or contextual use (González-Jover 2006, 224).

3. Hypothesis

My general hypothesis is that the source language influences the outcome of the translation process. Namely, certain specific features of SL texts which are uncom-

mon for TL texts can potentially appear in the translation, thus producing TT not consistent with TL norms. This interference might have certain universal features (i.e. features which are consistent across the majority of SL texts and their translations) as well as local features (i.e. those resulting from the given translator's affinity for transferring certain SL feature(s) into TL).

To support my hypothesis I conduct a case study of narrative verbs in English-Polish translations of fiction. By comparing data from a number of corpora (as specified in section 4.) I show how three translators (two of whom worked with the same three texts) handled the translation of the English verb *say* into its closest Polish equivalent *powiedzieć* 'say'. I underscore the statistical differences in their approaches and examine the patterns they followed.

4. Research material

To conduct this research it was necessary to create a number of corpora. I obtained samples at my own discretion (see Database for the titles) and removed the paratextual elements except for some of the footnotes (in Pratchett's novels these constitute an integral part of the texts). The corpora adhere to the statistical norms set forth by Sinclair and Biber. The *Polish Fiction Corpus* is large enough to examine syntactic patterns (Biber 1993, 252–255) whereas the parallel corpora are sampled proportionally (Sinclair 2005, 4).

The core of my research material consists of two large corpora, the *Cholewa Corpus* and the *Grupińska Corpus* (approx. 140,000 words each), obtained from the parallel translations of three books by Terry Pratchett. Each corpus contains the Polish translations of the same set of books by one translator. The translators are Piotr Cholewa and Dorota Malinowska-Grupińska, and the titles of the source texts are: *The Wee-Free Men*, *A Hat Full of Sky* and *Amazing Maurice and His Educated Rodents*. All the six translations were published between 2003 and 2011. The samples used in the creation of the corpora are fully parallel (so that every given sample X in the *Cholewa Corpus* has its directly equivalent sample Y in the *Grupińska Corpus*).

The *Pratchett Corpus* (approx. 135,000 words) was obtained from the original books by Terry Pratchett written in English. This corpus represents source texts and again contains samples fully parallel to those from the *Cholewa Corpus* and the *Grupińska Corpus*. Its *raison d'être* is to highlight the relationship between the source texts and the target texts allowing me to show the contrast between the two translations.

The *Polish Fiction Corpus* (approx. 1 million words) consists of large samples of 15 Polish fiction books written between 1980 and 2015 by various authors (see Database). This corpus is used as a general point of reference as to what a model native Polish fiction text should be like. Due to possibly large variations in the

writers' style, I make no claims solely on the basis of this corpus. This corpus does, however, represent consistent general trends for the language of Polish fiction and is therefore applicable when dealing with extreme deviations in the data.

Two support corpora (approx. 100,000 words each) were built to further exemplify the claims made on the basis of the core corpora. These were built from parallel samples of the fiction novel *Lies of Locke Lamora* written in English and its translation into Polish.

5. Research methodology

I use WordSmith Tools 6.0 to process the corpora into statistical data. WordSmith allows me to obtain raw data and run a number of tests. Initially I look for statistical peculiarities in the corpora, such as a substandard distribution of lexemes, unusual collocations, statistical use of grammatical words or striking lexical similarities between the source text and its translation. These can be discovered by a number of tests, as specified below.

When comparing parallel corpora (such as the *Grupińska Corpus* and the *Cholewa Corpus* which comprise translations of the same STs) I use statistical parameters such as STTR (standardised ratio between unique words in the given text and all the words in that text) as an indicator of differences on the syntactic level. High STTR discrepancy in the parallel translations of the same texts means that other qualities of the texts are likely to be diverse as well.

TTR (Type-Token Ratio) is a figure obtained by dividing the number of tokens (the sum of all the words in the given corpus) by the number of types (all unique instances of words) (Malvern and Richards 2000a, 2). Therefore, higher TTR means more varied vocabulary in the given corpus. To avoid the influence of texts' word counts on the results, the corpora are standardised, i.e. divided into 10,000 or 1,000 word samples, with TTR calculated for each such sample (Malvern and Richards 2000b, 85). The results for each sample are then summed and divided by the number of samples. Thus, STTR (Standardised Type-Token Ratio) is obtained.

To further explore particular differences between the texts, I use keyness tests. This test compares two corpora and lists the biggest discrepancies in vocabulary use and distribution by assigning the given word a keyness factor; the higher keyness, the greater the differences. For instance, if the word *say* is used two times more often in one translation than in the other, its keyness factor will be extremely high.

The keyness test has another application; for corpora representative of a given genre (containing a large quantity of texts from that genre), the patterns of vocabulary use and distribution for an average text in that genre are revealed. Though these are mostly trends and particular texts may deviate from them, many general features are very consistent across all texts in that genre (Scott 2010, 1). Therefore,

keyness tests prove particularly useful when comparing e.g. native Polish fiction and translated fiction.

Finally, when the texts are analysed with computer software, I look at the features that interest me (in this case, the narration). I carry out a comprehensive comparative analysis of the given feature across the corpora and examine the results. Having discovered a potential correlation between the number of occurrences of the English *say* in the English source texts and an unusually high number of occurrences of *powiedzieć* in the Polish target texts I decided to investigate this issue.

6. Analysis of the data

Even though STTR alone does not provide any particular information regarding the corpora, it is a valuable indicator of the differences between them. There is a visible discrepancy between both translation corpora and an average Polish fiction text (see Figure 1). It is important to note that in the *Polish Fiction Corpus* a vast majority of the books have STTR of 40 and above, and STTR was below 40 only in the case of two books. Hence, the first observation is that the translated texts may be characterised by a generally lower variety in their vocabulary.

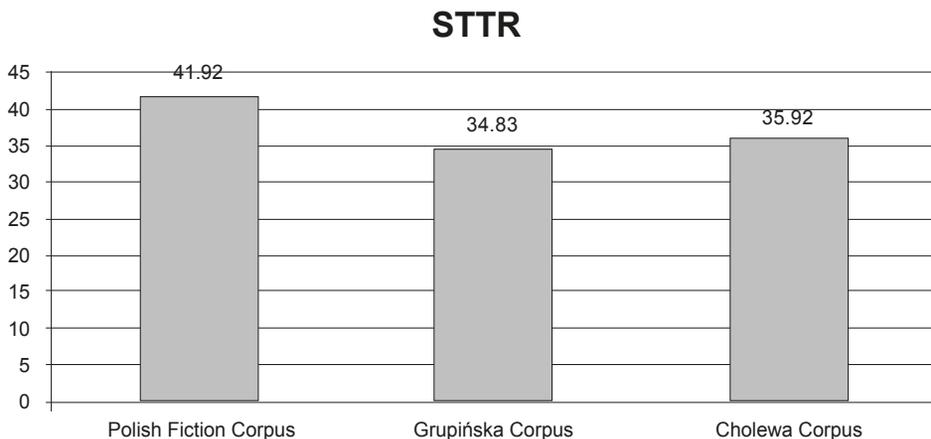


Fig. 1. STTR in the translations of the same three texts by different translators and in the corpus of native Polish fiction

The second comparison, however, is certainly more far-reaching. STTR in both translations of Pratchett's books (the *Grupańska Corpus* and the *Cholewa Corpus*) differs by 1.09%; though the difference may seem rather minute, it does regard translations of exactly the same source texts. This observation means that in Cholewa's translations the vocabulary is potentially far more diverse.

A further examination of the corpora revealed a number of differences between the two translation corpora. These involve the distribution of grammatical words, the use of collocations, the patterns in conjunct use and many others. When examining lexical verb use, however, I discovered a vast discrepancy in narrative verb use. This discrepancy appeared to be even more significant when I contrasted the translation corpora with the native Polish fiction corpus.

Narration, and more specifically the accurate use of narrative verbs, is a very important element of fiction texts as it introduces additional, specific information regarding characters' behaviour into the dialogues. Both Polish and English have a vast inventory of narrative verbs, which they use with varying frequency. Moreover, the equivalents for most English narrative verbs exist in Polish.

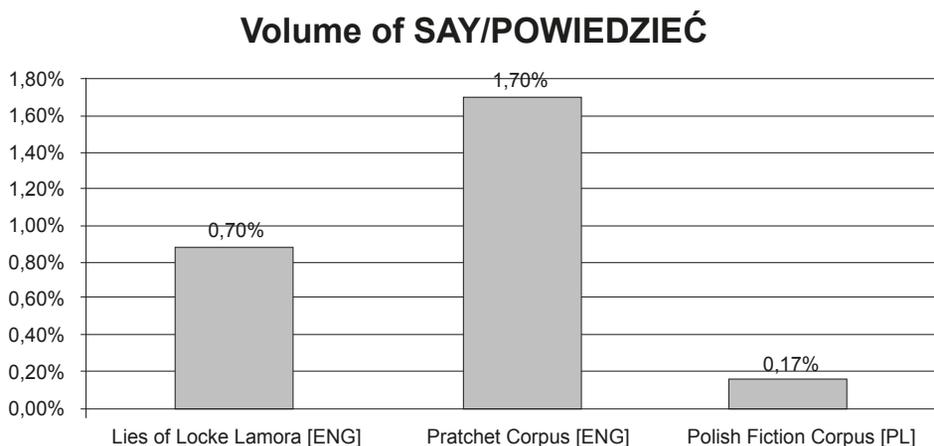


Fig. 2. The volume of the forms of *say* in the three texts by Pratchett, the text by Lynch and the volume of the forms of *powiedzieć* in the *Polish Fiction Corpus*

As exemplified in Figure 2, however, English has a strong preference for forms of *x said* or *said x*. Even though other narrative verbs are in use, the forms of the verb *say* are highly prevalent. For the native Polish texts the frequency of the direct equivalent of *say*, i.e. *powiedzieć*, is far lower. Often no narrative verb is used at all or various other synonymous words are used instead. This variance of *say* being replaced by more specific verbs does obviously occur in English fiction texts too but to a far lesser extent.

Data comparison (Figure 2) clearly demonstrates the statistical difference in the use of *say* and *powiedzieć* between Polish and English. In the *Polish Fiction Corpus* the forms of the word *powiedzieć* occupy 0.17% of the text whereas in English the figure reaches 1.7% in a dialogue-heavy corpus and 0.7% in a less dialogue-heavy fiction text. In the corpora I examined the difference between the Polish and English fiction texts in this regard is between 3x and 10x.

My aim is to examine how the abundance of *say* in the English source texts influences the use of its direct translation, namely the forms of *powiedzieć*, in the target texts. The diagram below (Figure 3) demonstrates the relationship between the source texts (tagged as ENG) and their translations into Polish (tagged as TRANS). It is worth noting that the texts from the *Grupińska Corpus* and the *Cholewa Corpus* are the translations of the texts in the *Pratchett Corpus*. Moreover, all the three corpora were built in a parallel fashion.

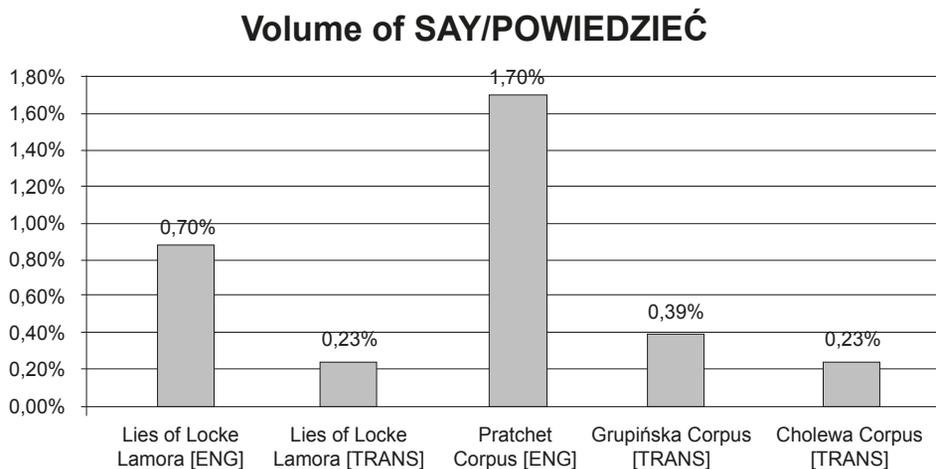


Fig. 3. The volume of the *say/powiedzieć* forms in the source texts and in their translations into Polish

This diagram captures a number of interesting dependencies (or, perhaps, non-dependencies). Firstly, the volume of *powiedzieć* in the *Cholewa Corpus* and in the *Lies of Locke Lamora [TRANS] Corpus* is very close to that of the *Polish Fiction Corpus*. Importantly, though, there is a visible discrepancy between the use of *say* in the source texts of those two translation corpora, with the *Pratchett Corpus* containing more than two times more forms of *say* than the *Lies of Locke Lamora Corpus*. Secondly, there is a visible difference between the *Grupińska Corpus* and the *Cholewa Corpus*, whereby *Grupińska's* use of the *powiedzieć* forms almost doubles *Cholewa's* (which is the main concern of this paper).

Two parallel corpora (the *Grupińska Corpus* and the *Cholewa Corpus*) are my main sources of data. Since they use the translations of the same source texts, one would assume they are rather similar on the lexical level, with small variance generated by idiolectal differences between the translators. According to my hypothesis, the translations should, however, display certain qualities of their English counterparts.

Volume of POWIEDZIEĆ

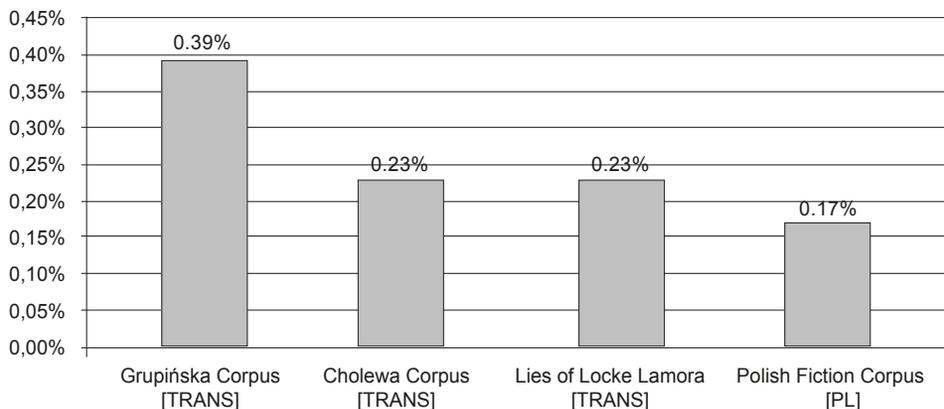


Fig. 4. The volume of the *powiedzieć* ‘say’ forms in the texts translated into Polish and in the native Polish text corpus

The distribution of the verb *powiedzieć* in the *Polish Fiction Corpus* indicates that the fiction texts written in Polish will on average have approximately 0.17% of the *powiedzieć* forms in their volume. Based on the data from individual books from the corpus I assume 0.15–0.25% to be a standard volume for *powiedzieć* for a native Polish fiction text. It is worth noting that even the specific texts rich in dialogues would nevertheless rarely exceed 0.20%.

Cholewa’s translations are in accord with the standards I set forth above. He attained a 0.23% volume of *powiedzieć* even when translating a set of books with an excessively high volume of *say* (for English fiction texts). Moreover, his translations appear to share a number of other characteristics with the texts of the *Polish Fiction Corpus* (including grammatical word distribution and syntactic patterns). I therefore claim that the fact that Cholewa translated from English did not influence the narrative verbs in his translation to a statistically relevant degree.

In contrast with the above, Grupańska’s translations doubly exceed the standards of the *Polish Fiction Corpus* with regard to the distribution of the *powiedzieć* forms. Indeed, there appears to be a correlation between Grupańska’s translations and the density of the *say* forms in the source text. Even though 0.39% of the TTs is nowhere near 1.7% of the STs, the figure is significant enough and unnatural for the standards of a Polish text I assumed. The data become particularly noticeable when the *Grupańska Corpus* is further contrasted with the *Cholewa Corpus* (Figure 4). Even though Cholewa managed to retain a high variety of narrative verbs in his translation (only 0.23% of the *powiedzieć* forms) of the same set of texts, Grupańska visibly failed to do so.

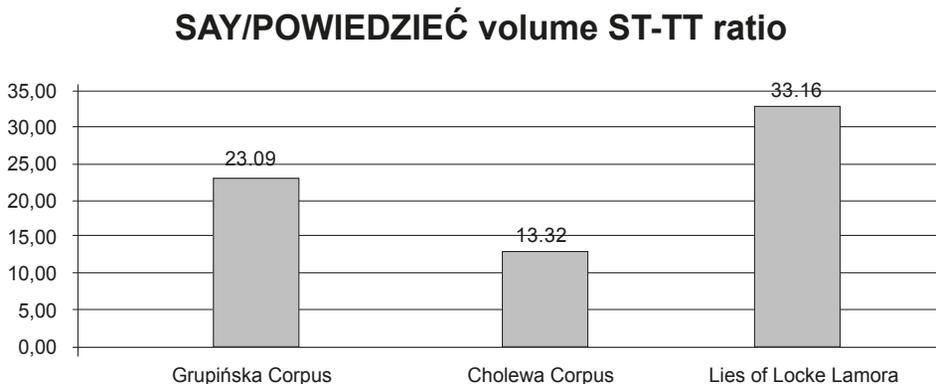


Fig. 5. The ratio of the volume of the *say/powiedzieć* forms between ST and TT

The final element of my analysis is a comparative account of the relationships between the number of the *say* forms in the ST corpus and the number of the *powiedzieć* forms in the TT corpus (Figure 5). The higher ratio in the diagram indicates a closer relationship between ST and TT in that regard. Firstly, as explained before, Grupańska's translations are strongly related to their STs whereas Cholewa has retained the balance of narrative verbs appropriate for Polish. Surprisingly, however, the translator of *Lies of Locke Lamora* does also show a strong tendency to use the literal translation of *say*.

However, this noticeable correlation for the narrative verb use in the translation of *Lies of Locke Lamora* and its ST does partially stem from the fact that the ST uses a particularly rich inventory of narrative verbs. It is therefore possible that due to the more varied synonyms of *say* the translator opted for literal translation for it was not necessary to overcomplicate the dialogues. This final piece of data does therefore remain inconclusive.

7. Conclusions

My data indicate that the high frequency of the narrative verb *powiedzieć* in some English-Polish translations is likely to be caused by the interference from the source text. When contrasting the distribution of *powiedzieć* in the *Cholewa Corpus* with that in the *Grupańska Corpus*, the relationship between ST-TT direct translation of *say* becomes particularly apparent. While the distribution of *powiedzieć* in the *Cholewa Corpus* quite closely resembles that of a native Polish fiction text (0.06% deviation), the *Grupańska Corpus* exceeds the use of this expression by 0.22%. The comparison of ST-TT ratio (Figure 5) shows that the connection between the ST use of *say* and the TT use of *powiedzieć* was weak (13.32) for the *Cholewa Corpus* and almost twice stronger for the *Grupańska Corpus* (23.09). This relationship is

a strong indicator that interference of the forms of *say* has occurred in the work of Grupińska. Even though the *Cholewa corpus* has exceeded the average use of *powiedzieć* in the *Polish Fiction Corpus* by 0.06%, this variation is not significant enough to support any claims.

The relationship in the use of *say* and *powiedzieć* between the ST and the TT for the *Lies of Locke Lamora Corpus* (Figure 5) is also very strong (33.16, so stronger than in the *Grupińska Corpus*). It is important to note, however, that the volume of *powiedzieć* itself remains within the standards of a Polish fiction text. I therefore assume that this relationship was not caused by interference but rather by the fact that the ST already had a broad range of synonyms of *say* in use and therefore there was no need for the translator to re-adjust the TT.

In my future papers I aim to continue comparing Polish TTs with their English STs. While conducting research on narrative verbs I discovered a number of other peculiarities in my corpora, though I was not able to measure their impact yet. I also aim to continue building my *Polish Fiction Corpus* and *English-Polish Translation Corpus* so that they contain data strong enough for a more elaborate statistical analysis.

References

(a) Database

Pratchett Corpus

Pratchett, Terry. 2001. *The Amazing Maurice and his Educated Rodents*. London: Doubleday.

Pratchett, Terry. 2003. *The Wee Free Men*. London: Doubleday.

Pratchett, Terry. 2004. *A Hat Full of Sky*. London: Doubleday.

Cholewa Corpus

Pratchett, Terry. 2010. *Wolni Ciut Ludzie*. Trans. Piotr Cholewa. Warszawa: Prószyński i S-ka.

Pratchett, Terry. 2011. *Kapelusz pelennieba*. Trans. Piotr Cholewa. Warszawa: Prószyński i S-ka.

Pratchett, Terry. 2011. *Zadziwiający Maurycy i jego edukowane gryzonie*. Trans. Piotr Cholewa. Warszawa: Prószyński i S-ka.

Grupińska Corpus

Pratchett, Terry. 2004. *Zadziwiający Maurycy i jego edukowane szczury*. Trans. Dorota Malinowska-Grupińska. Warszawa: Prószyński i S-ka.

Pratchett, Terry. 2005. *Kapelusz pelen nieba*. Trans. Dorota Malinowska-Grupińska. Warszawa: Prószyński i S-ka.

Pratchett, Terry. 2005. *Wolni Ciutludzie*. Trans. Dorota Malinowska-Grupińska. Warszawa: Prószyński i S-ka.

Lies of Locke Lamora Corpora

Lynch, Scott. 2006. *Lies of Locke Lamora*. London: Gollancz.

Lynch, Scott. 2013. *Kłamstwa Locke'a Lamory*. Trans. Wojciech Szypuła, and Małgorzata Strzelec. Warszawa: Mag.

Polish Native Corpus

Brzezińska, Anna, and Grzegorz Wiśniewski. 2007. *Wielka wojna*. Warszawa: Runa.

Dukaj, Jacek. 1993. *Irrehaare*. Kraków: Wydawnictwo Literackie.

Hugo-Bader, Jacek. 2010. *W rajskiej dolinie wśród zielsk*. Wołowiec: Czarne.

Karpowicz, Ignacy. 2006. *Niehalo*. Kraków: Wydawnictwo Literackie.

Karpowicz, Ignacy. 2010. *Balladyny i romanse*. Kraków: Wydawnictwo Literackie.

Karpowicz, Ignacy. 2013. *Ości*. Kraków: Wydawnictwo Literackie.

Kres, Feliks. 1992. *Król Bezmiarów*. Warszawa: Aurora.

Krajewski, Marek. 2010. *Erynie*. Kraków: Znak.

Krajewski Marek. 2011. *Liczby Charona*. Kraków: Znak.

Krajewski, Marek. 2012. *Rzeki Hadesu*. Kraków: Znak.

Lem, Stanisław. 1961. *Pamiętnik znaleziony w wannie*. Kraków: Literackie.

Myśliwski, Wiesław. 2013. *Ostatnie rozdanie*. Kraków: Znak.

Oramus, Marek. 1985. *Arsenał*. Warszawa: Iskry.

Szostak, Wit. 2004. *Poszarpane granie*. Warszawa: Runa.

Tokarczuk, Olga. 1998. *Dom dzienny, dom nocny*. Kraków: Wydawnictwo Literackie.

Varga, Krzysztof. 2014. *Czardasz z mangalicą*. Wołowiec: Czarne.

Zajdel, Janusz. 1980. *Cylinder Van Troffa*. Warszawa: Czytelnik.

Zajdel, Janusz. 1984. *Paradyzja*. Warszawa: Iskry.

(b) Special studies

Baker, Mona. 1992. *In Other Words*. London: Routledge.

Baker, Mona. 2001. "Towards a Methodology for Investigating the Style of a Literary Translator." *Target* 12.2: 241–266.

Biber, Douglas. 1993. "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8.4: 243–257.

Cook, Vivian. 2003. "Multi-competence: Black Hole or Wormhole for Second Language Acquisition Research." *Understanding Second Language Process*. Ed. Zhaohong Han. Multilingual Matters. 16–26.

Crystal, David. 2003. *A Dictionary of Linguistics and Phonetics*. New York: Wiley.

Fokin, Sergiy. 2015. "TTR Changes in Different Directions of Translation." <http://translationjournal.net/journal/63ttr.htm>.

- González-Jover, Adelina Gómez. 2006. "Meaning and Anisomorphism in Modern Lexicography." *Terminology* 12.2: 215–234.
- Malvern, David, and Bryan Richards. 2000a. *Measuring Vocabulary Richness in Teenage Learners of French*. <http://www.leeds.ac.uk/educol/documents/00001541.htm>
- Malvern, David, and Bryan Richards. 2000b. "Validation of a New Measure of Lexical Diversity." *From Sound to Sentence: Studies on First Language Acquisition*. Ed. Mieke Beers. Groningen: Centre for Language and Cognition. 81–96.
- Mauranen, Anna. 2004. "Corpora, Universals and Interference." *Translation Universals: Do They Exist?* Ed. Anna Mauranen, and Pekka Kujamäki. Amsterdam: John Benjamins Publishing. 65–82.
- Nilsson, Per-Ola. 2004. "Translation-specific Lexicogrammar? Characteristic Lexical and Collocational Patterning in Swedish Texts Translated from English." *Translation Universals: Do They Exist?* Ed. Anna Mauranen, and Pekka Kujamäki. Amsterdam: John Benjamins Publishing. 129–145.
- Nord, Christine. [1997] 2014. *Translation as a Purposeful Activity. Functionalist Approaches Explained*. Manchester: Saint Jerome Publishing.
- Pym, Anthony. 1992. *Translation and Text Transfer*. Tarragona: Intercultural Studies Group.
- Scott, Mike. 2010. *WordSmith Tools Manual*. http://www.lexically.net/downloads/version6/HTML/keywords_calculate_info.htm
- Sinclair, John. 2005. "Corpus and Text-basic Principles." *Developing Linguistic Corpora: A Guide to Good Practice*. Ed. Martin Wynne. Oxford: Oxbow Books. 1–16.