λ

EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-9 September 2021
Virtual

www.euralex2020.gr

Speech
Etymology
Idioms
Glossary
NLP
Lemma
Meaning
Corpora
Word
Lexicon
Dictionary
Definition
Pronunciation
Headword
Entry
Lexicology
Dictionary Use
Lexical Resources

2021 Edition

Charting A Landscape of Loans. An e-Lexicographical Project on German Lexical Borrowings in Polish Dialects

Meyer P., Hentschel G.

[1] *Fryske Akademy, Leeuwarden (NL)*
[2] *Universiteit Utrecht, Utrecht*
edrenth@fryske-akademy.nl; hsijens@fryske-akademy.nl; hvandevelde@fryske-akademy.nl

Abstract

This paper reports on an ongoing international project of compiling a freely accessible online Dictionary of German Loans in Polish Dialects. The dictionary will be the first comprehensive lexicographic compendium of its kind, serving as a complement to existing resources on German lexical loans in the literary or standard language. The empirical results obtained in the project will shed new light on the distribution of German loanwords among different dialects, also in comparison to the well-documented situation in written Polish. The dictionary will have a strong focus on the dialectal distribution of Polish dialectal variants for a given German etymon, accessible through interactive cartographic representations and corresponding search options. The editorial process is realized with dedicated collaborative web tools. The new resource will be published as an integrated part of an online information system for German lexical borrowings in other languages, the *Lehnwortportal Deutsch*, and is therefore highly cross-linked with other loanword dictionaries on Polish as well as Slavic and further European languages.

Keywords: lexical borrowings; Polish dialectology; dialect lexicography; XML database

1 Introduction and Lexicographical Background

This paper reports on an ongoing project of compiling a freely accessible online Dictionary of German Loans in Polish Dialects (henceforth, DGlPd) funded by the German Research Foundation (DFG) and jointly carried out by the Institute of Slavic Studies at the University of Oldenburg and the Leibniz Institute for the German Language (IDS), in cooperation with the Institute of the Polish Language of the Polish Academy of Sciences, Kraków (IJP PAN). The project started in April of 2019 and will end in 2022/23.

Scientific research on German loans in Polish (and other Slavic languages) started in the 19th century. Older studies of the phenomenon did not have the opportunity to take advantage of the enormous progress of historical and dialectal lexicography in Poland (and other Slavic speaking countries, cf. SALDAS for bibliographical references) in the decades after World War II up until today. The DGlPd will be the first comprehensive lexicographic compendium of its kind, serving as a complement to the Dictionary of German Loans in the Polish Written and Standard Language (WDLP).

The empirical results obtained in the project will shed new light on the distribution of German loanwords among different dialects, also in comparison to the well-documented situation in written Polish as documented in the WDLP (cf. Hentschel 2009; 2010). One of the general questions to be asked is to what extent the transfer of German loans was mediated by dialects and to what extent dialects took over loans from the written language. A small quantitative pilot study was carried out in preparation for the project, based on entries in the SGP for the initial letters <A> and <F>, not counting derived forms, totalling to 529 elements. The study indicates that Polish "core dialects" such as those of Lesser Poland, Greater Poland and Mazovia that had almost not at all, or only in some parts, been under long-term German (speaking) rule (Greater Poland, Mazowia) or just for a comparatively limited time (Lesser Poland from the end of the 18th century to World War I, though) show significantly more overlap in German loans with written Polish than dialectal areas with a centuries-long direct contact situation, including bilingualism in large parts of the population. Within the latter group, e.g. Silesian and Kashubian (see below for the linguistic status of the latter) have borrowed extensively from German but hardly had an impact on the development of Literary Polish, while core dialects definitely had. In late medieval until early modern times a broad scale direct contact of Poles with German migrants to Poland was a main source of lexical transfer in large parts of Poland. From the 19th century to the end of World War II, such a situation was given only in Silesia, Kashubia and parts of East Prussia. Nevertheless, there is a considerable amount of German loans dating back to the 19th century, when large parts of Poland (viewed in its contemporary boarders) were under Prussian or Habsburg rule. In the DGlPd, special attention will thus be paid to the amount of overlap of core dialects and non-core dialects in German loans on the one hand and to the overlap of different dialects with written Polish.

There are two theoretical aspects of lexical borrowings that have recently received more attention in the literature and that are challenges to our investigation. The first one concerns the discussion in contact linguistics on code switching and code mixing (cf. Muysken 2000) concerning insertional code switching of single lexical items and so-called spontaneous or nonce borrowings (cf. for example Poplack & Dion 2012), given that adaptation of transferred elements in a recipient language is not always a reliable cue to the problem. In non-core dialects, the speakers of which are at least to some extent bilingual, with German as the dominating "H-variety", the mere presence of a lexical item in the standard dictionary of Polish Dialects, SGP, is no guarantee for acceptance of the item by the speech community as a whole. Here only the number and distribution of witnesses can roughly mirror the degree of acceptance. The second aspect is an issue in

typological investigations of borrowing (cf. Haspelmath & Tadmor 2009). Following the lines of Haspelmath (2009), an attempt will be made in the DGlPd to differentiate between "insertions" (note: not in the sense of "insertional code switching / mixing") into the lexicon (roughly: cultural borrowings, to name new things or concepts), "replacements" (roughly: borrowings replacing an older, typically native Polish word), and – as we would like to add – constellations of "coexistence" involving both the new loanword and an older word, with some type of reorganization of the denotational scope of the borrowed and the native (or at least older) word.

In what follows, we will first present the data sources and scope of the lexicographical project (section 2) before outlining the editorial process (section 3) and the IT tools used in this process (section 4). Section 5 gives an overview of the planned entry microstructure, while relevant details of the online presentation are covered in section 6. #wrapup/conclusion in section 7.

2 Source and Scope of Data

The central, though not exclusive, source of data for the project is the Dictionary of Polish Dialects (SGP), which is being compiled in the Department of Dialectology of the IJP PAN. Begun in 1982, the SGP now counts nine released volumes, up to the lemma *hyżki*. This means that most of the SGP data, roughly three quarters, has to be taken from the card index to the SGP. It should be noted that the SGP does not give any information on the origin of the words described.

In accordance with the conception of the SGP, all traditionally acknowledged dialects of Polish will be taken into account in the dictionary: Silesian, Greater Polish, Lesser Polish, Mazovian, Northern and Southern Kresy, as well as Kashubian, which had been regarded as a Polish dialect in Poland for a long time. This point of view can be found in the two-volume compendium *Języki indoeuropejskie – The Indo-European languages* (Bednarczuk 1986-88), authored by highly acknowledged representatives of Polish linguistics of that time (vol. II, pp. 919f). In 2005, Kashubian was officially given the status of a regional language. Since the Polish language law explicitly rules out that a dialect of the official language of the state (i.e. Polish) can be given the status of a regional language, this means that Kashubian is no longer regarded as a dialect of Polish (USTAWA §19).

Similar to the WDLP, only loanwords that are not themselves loans in German (coming from Italian, French, Latin, to name the most important ones) will be considered. Given the considerable number of German loans with German(ic) etymology in Polish dialects (we estimate about 5,000 elements, without derived forms, compared to some 2,500 in Written Polish according to WDLP), the number of loans from other ("Western") languages, which have possibly been mediated by German, probably has a similar order of magnitude, which would surpass the practical possibilities of a 3-year project. There is, however, a more qualitative justification for limiting the scope of loanwords. For each possibly mediated loan in Polish it would be necessary to determine whether the loan has in fact been mediated by German or whether it is a direct loan from some other language – a question that, incidentally, in many cases would have to be left open anyway. In the latter case, we would expect these words to have been borrowed into Standard / Cultural Polish and only later taken over by dialects as direct language contact and bilingualism was restricted to the elites. Until now there has been no systematic and comprehensive investigation of such loans in Standard / Cultural Polish, with studies like Walsleben (1997) covering only chronological or thematic parts of the lexicon. In other words, decisive prerequisites for the inclusion of such, possibly mediated, elements are currently lacking, whereas German loans with German(ic) etymology in Standard / Cultural Polish have been fixed and documented by the WDLP.

In one respect, the DGlPd will not follow the conception of the WDLP. The latter does not comprise German loans, even with Germanic etymology, if they were mediated by Czech to Cultural / Written Polish before the beginning of the 17th century (the Czech impact on Polish vanished afterwards). Our different approach is motivated by the possibility that a German loan, though mediated by Czech to Written Polish, has been taken over into some Polish dialects (e.g. Silesian) directly from German. There are, for example, some language islands of dialectal Czech in Upper Silesia, which stand in direct contact with varieties of Polish. If a German loan fulfils the condition of German(ic) etymology, then it will be included in the DGlPd. This means: (i) A German etymon under consideration must not be described as being transferred into German from other languages, mainly Latin and Romance languages such as Italian and French. (ii) In order to categorize a word in a receiving language as being transferred from a donor language the corresponding word in the latter should exhibit a substantial affinity to the potential source word: (a) in expression, taking into account known interlingual sound substitutions, here between German and Polish, as well as intralingual sound relations between dialects of the receiving language and universal phonological processes and (b) in meaning, taking into account semantic processes such as the narrowing of meanings from source word to loan as well as metaphoric and metonymic meanings shifts. Last but not least, for classifying a word as a loan from some other language socio-historical plausibility is an important aspect.

3 Compilation Process

The compilation of the DGlPd proceeds roughly as follows. As a first step, proto-entries are created that each connect phonetically similar Polish dialectal word forms, taken from a comprehensive Polish word index of the SGP (cf. section 4), to a small candidate set of candidate German etyma (due to a considerable amount of phonetic variation of Polish dialectal words, their relation to possible German etyma is in many cases rather opaque). After that, detailed information on the Polish words, in particular on their attested meanings and dialectal distribution, obtained from the published volumes and the card index of the SGP, is inserted into these proto-entries. As a last step, the final entries are constructed by splitting and merging the proto-entries according to an etymological assessment guided by the attested word senses, adding an essential commentary on etymology and word history in general.

As indicated above, roughly three quarters of the material needed for the DGlPd still only exists in the form of a huge number of cards that, for the time being, have been sorted only on the basis of unlemmatized expressions. This means that

the project partners at the IJP PAN in Kraków must manually select all possibly relevant cards for each loan preliminarily selected on the basis of the alphabetical index. This involves (i) checking the card index for expression variants (and there can be many at very different alphabetical positions, cf. Polish dialectal equivalents to German *Hundsfott* as noted in SGP: *hunctwot, huncwót, huncwód, huncót, huncwat, hunzwot, hunsot, huńcot, huńcót, hucwont, hucwont, hacnont, hyncwant, hicnond, uncwot, wuncwot*); (ii) classifying them for different meanings; (iii) fixing the areal distribution of meanings, expression variants and derived words; (iv) selecting citations for meanings and, where possible, for expression variants. With an estimated 5,000 entries of the DGlPd (not counting derivative forms), the card index must be consulted for almost 4,000 entries, meaning that tens of thousands of cards will have to be checked to collect the information needed. As the card index has been collated over the course of many decades, by a large number of people, only the colleagues working at the location in Kraków have the competence to interpret the heterogeneous transcriptions and can thus guarantee the reliability of the notation of the expressions and the citations in the DGlPd.

4   Tooling

For the very specific purposes and requirements of the project, an in-house web application with role-based user management has been developed for collaborative compilation and editorial work in Oldenburg and Kraków. Similar to the LeXmart dictionary development framework (Simões et al. 2019), the software uses an XML database management system (in our case, BaseX, https://basex.org) as its storage backend and features the JavaScript library Xonomy (also used in the well-known Lexonomy system, cf. Měchura 2017) as its browser-based XML entry editing component for the fully bilingual German/Polish user interface. The bare XML editor functionality is supplemented by a growing range of project-specific tools.

- Words and their attested senses must be "localized" according to a scheme that differentiates roughly 350 dialects, subdialects and townships. Frequently, a large number of such localisations has to be selected for a single word or word sense. A dedicated input system speeds up this process with autocomplete functionality, multiple selection within dialect groups, and visual feedback as to whether localisations for words and their word senses match. No manual XML editing is required.
- The 'proto-entries' often comprise a large number of word forms and word senses. In the editing process, the senses have to be assigned to the word forms they are attested for. A generic linking mechanism based on universally unique identifiers is used to track the 'cross-referencing' between words and their senses even across entries. Users edit the presence or absence of connections between form and meaning in a convenient matrix representation, again without having to bother with XML or even UUID minutiae. The general mechanism, which does not encode the links within the entries but as separate XML elements, can later be used to e.g. associate Polish words with their etyma.
- For the final editing phase, a comprehensive tool to split and merge entries, reallocate words and senses to other entries, reclassify expression variants as derivatives and vice versa, etc. is needed. Care has to be taken not to inadvertently sever the association between a word reallocated to a new entry and its senses. The number of proto-entries involved in a single consolidation process can be large, implying that the tool must abstract away from the XML source to give the lexicographer the overview needed.
- Reporting of lexicographical activity and results is done through a powerful XQuery-based interface. During editing, time-stamped snapshots of the current entry are stored in the database automatically in regular intervals, which means that the complete project history can be accessed in database queries such that arbitrary kinds of editing processes can be tracked easily. For each (proto-)entry, all of its editing snapshots are only a mouse click away.
- Entries can be accessed and looked up via their German etyma and the Polish words contained in them. In particular, the entire Polish word index of the SGP (approx. 230.000 words, including phonological and morphological variants and derivatives) is available in the browser, as well as the full XML datasets of two other lexicographical projects related to German loanwords in Polish, viz. the WDLP and the WDLT.
- The XML source code for large entries can quickly get too long for convenient editing. An XSLT-based entry preview function allows lexicographers to easily "jump" from any position in the preview to the corresponding position in the XML editor, just by clicking on the position.
- There is a simple, extensible system of macros to enter special characters without having to change (virtual) keyboards.

Figure 1 shows a screenshot of the basic editor system. The tabs on the left-hand side provide immediate access to the dictionary's (proto-) entries, the Polish word index, and the two Polish dictionaries WDLP and WDLT cross-referenced in the DGlPd entries. Figure 2 shows a screenshot of the localisation editing dialog for the expression variant *klapa*, with a multiple select input option for localisations within a dialect and colour feedback as to whether, in the case shown, localisations provided for the word also appear in the word senses connected to it and vice versa.

Figure 1: Basic XML editor screen.

Figure 2: Localisation editor.

5   Entry Microstructure

The estimated 5000 entries of the DGlPd group Polish dialectal words together with their word senses according to their common German etymon in a nest structure. The currently planned main lexicographical indications are enumerated below. A detailed lexicographic protocol will depend on the outcome of the compilation process for the 'proto-entries' as explained above in section 3.

- The German **etymon** is provided with grammatical information and relevant word senses. Frequently, multiple etyma, usually with close diasystemic ties to each other, must be given.
- All attested Polish **dialectal variants** for a given German etymon are listed, together with grammatical information. The depth of the differentiation of dialectal "phonic" – i.e. phonological or phonetic – expression variation will be the same as in the SGP. The variants will be offered in a normalized graphic form, again based on the rules applied in the SGP. Each expression variant will receive a rough classification as to its degree of deviation from the expression of the German etymon, namely (i) no or minor deviation, (ii) medium deviation, or (iii) strong deviation. An "approximation algorithm" will be used for the classification, conceptually similar to an (informal) variant of the Levenshtein distance and based on the number of phonological segments affected by the transfer of expression material, the number and the quality of the phonological processes that could be seen behind phonic substitutions and the fact that certain observed phonic substitutions cannot be explained by known types of phonological processes (cf. Stachowski 2011). Such a classification will at least have a heuristic value and will, as one of many search options, offer valuable information for linguistic research on mechanisms of material transfer of expressions.
- The entries will provide the **word senses** attested for the expression variants and map them to a semantic field classification roughly along the lines of Haspelmath/Tadmor 2009.
- **Primary derivatives** will be listed with their forms and word senses.
- Information will be provided on **equivalents in Eastern Yiddish**, taking advantage of lexicographic sources such as as Stutchkoff (1950) and Astravuch (2008).
- In a similar vein, the entries will provide information on **equivalents in (Old) Czech**, mainly based on Newerkla's comprehensive study on German loans in Czech and Slovak (Newerkla 2011).
- Each entry features a **textual commentary** with an assessment of the history of the loanwords.

Word variants, derivatives, and word senses are each assigned "localisations" (i.e. areas of dialectal attestation) which may range from whole dialectal and subdialectal groups down to the level of individual counties (*powiaty*).

For each word sense, representative citations for select localisations will be offered. It has to be taken into consideration that in contrast to dictionaries based on written texts, the fixation of a word in a concrete dialect or subdialect, i.e. in the corresponding entry of the card index, sometimes only exists in the form of "expression E (with meaning M) in village / region V".) In such cases no citations can be offered.

6   Online Presentation

The DGlPd online entry presentation will be fully bilingual, which means that even word sense definitions, paraphrases, textual commentary, etc. will be offered in both German and Polish.

The entries will feature interactive cartographic representations that visualize the available fine-grained localisations, aggregating the data on roughly the level of a traditional map of Polish dialect space, with the traditional dialect areas differentiated in greater detail based on the time periods of German (speaking) rule. There will be at least one global map for each entry, but depending on the degree of complexity of formal variation, polysemy and/or derived forms further maps on specific words or meanings may be necessary for reasons of transparent exposition.

All entries containing data from the as yet unpublished parts of the SGP will contain hyperlinks to digital images of the relevant cards – many of them hand-written – of the card SGP index.

The DGlPd will be integrated into the *Lehnwortportal Deutsch* (LWP; cf. Meyer 2013), an online publication platform for a growing number of dictionaries on German lexical borrowings in other languages. In the LWP, the lexicographical data is internally managed as a cross-dictionary network (technically, a graph database) of relations between word forms (etyma, loanwords, the corresponding meanings and expression variants, derivatives, etc.) of all included dictionaries. In a third-party project funded by the Fritz Thyssen Foundation, the LWP, which is hosted and maintained at the IDS, is currently undergoing a complete redesign with respect to underlying technologies and user interface design. The resulting system, publicly available in 2022, will be based on a refined, manually curated and autonomous graph-based abstraction layer on top of the 'native' dictionary data that allows users to find and navigate through lexical units and their relations to each other in real time in an interactive, cross-resource fashion (Meyer & Eppinger 2018). The revamped version will start with about ten newly added resources, most of them originally print dictionaries, covering many important European recipient languages such as English, French, Dutch, Czech, Slovak, and Hungarian. Even now, the LWP hosts digitally enhanced versions of two Polish-related resources, viz. the WDLP and the WDLT, soon to be complemented by an online version of a recently published frequency dictionary on German loans in the Silesian dialect of Polish (Hentschel, Tambor & Fekete 2021). Starting 2022, the most comprehensive study on German loans in Czech and Slovak (Newerkla 2011) will be part of the system. Somewhat later, a specialised dictionary on parallel lexical borrowings from German in Polish and in the East Slavic Languages (Belarusian, Ukrainian, Russian) will be available in the LWP (cf. Meyer 2015 for an early presentation). Similar to the DGlPd, this dictionary is based on a cooperation between the Institute of Slavic Studies at the University of Oldenburg and the Leibniz Institute for the German Language.

Several of the resources named here (WDLP; WDLT; Newerkla 2011) are systematically referenced in the DGlPd entries;

together with the mostly etymon-based cross-resource links in the LWP this will embed the DGlPd data in a large network of Polish-related information concerning Literary and dialectal Polish and important contact languages.

The newly designed search options will allow users to find information based both on general, cross-dictionary criteria and on dictionary-specific criteria. For the DGlPd, the latter will include the possibility to query the fine-grained localisation data as well as the highly specialised classifications, named in section 5, for dialectal variants and word senses. Users will be able to formulate arbitrarily complex queries (including, of course, Boolean operators, regular expressions etc., but also descriptions of multiple-word borrowing constellations) through a visual query builder (Meyer 2019) integrated seamlessly into the main page.

7   Conclusion

The ongoing lexicographical endeavour outlined in this paper aims to present the linguistic results of longstanding lexical borrowing processes from German into dialectal Polish almost literally as a complex landscape – to be digitally explored by experts and interested laypersons alike. Besides a more traditional, linear textual dictionary entry format, users are offered cartographic visualizations of the dialectal distribution of loanwords and their meanings. A dense network of cross-references to other resources on German loans in Polish and other, mainly European, languages, can be leveraged in advanced search queries that can take into account the relations between lexical units available in the LWP's graph database. Users may even get a glimpse of the lexicographical practice underlying the DGlPd's main source, looking at images of index cards collected during decades and now used for a publication the first time.

The DGlPd can be seen as a further step toward a full representation of lexical traces of German cultural and linguistic contact in Eastern Europe. Through its integration in a platform of interlinked multilingual resources it will remain open towards future lexicographical, dialectal, and historical research.

8   References

Astravuch, A. (2008). *Idyš-belaruski sloŭnik.* Minsk: Medysont.

Bednarczuk, L. (1986-1988) Języki indoeuropejskie. Vol. I / II. Warszawa: PWN.

Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In: M. Haspelmath, U. Tadmor (eds.) (2009), pp. 35-53.

Haspelmath, M., Tadmor, U. (eds.) (2009). Loanwords in the world's languages: A comparative handbook. Berlin: Mouton de Gruyter.

Hentschel, G. (2009). Intensität und Extensität deutsch-polnischer Sprachkontakte von den mittelalterlichen Anfängen bis ins 20. Jahrhundert am Beispiel deutscher Lehnwörter im Polnischen. In C. Stolz (eds.) *Unsere sprachlichen Nachbarn in Europa. Die Kontaktbeziehungen zwischen Deutsch und seinen Grenznachbarn.* Bochum: Brockmeyer, pp. 155-171.

Hentschel, G. (2010). *Zur Einführung [Introduction to the WDLP] online.* Accessed at: http://diglib.bis.uni-oldenburg.de/bis-verlag/wdlp/46704.html [15/04/2021].

Hentschel, G., Tambor, J. & Fekete, I. (2021). *Frequenzwörterbuch deutscher Lehnwörter im Schlesischen der Gegenwart. Mit Kommentaren zur Etymologie.* Oldenburg: BIS-Verlag.

Měchura, M. B. (2017). Introducing Lexonomy: an open-source dictionary writing and publishing system. In I. Kosem et al. (eds.) *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference, 19- 21 September 2017, Leiden, The Netherlands.* Leiden, The Netherlands. Brno, pp. 662-679.

Meyer, P. (2013). Ein Internetportal für deutsche Lehnwörter in slavischen Sprachen. Zugriffsstrukturen und Datenrepräsentation. In: S. Kempgen, M. Wingender, N. Franz, M. Jakiša (eds.): *Deutsche Beiträge zum 15. Internationalen Slavistenkongress.* Minsk, München: Otto Sagner, pp. 233-242.

Meyer, P. (2015). Aligning word senses and more: tools for creating interlinked resources in historical loanword lexicography. In K. Kallas, I. Kosem, S. Krek (eds.*) Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom. Ljubljana/Brighton.* Trojina: Institute for Applied Slovene Studies, pp. 198–210.

Meyer, P., Eppinger, M. (2018). fLexiCoGraph: Creating and Managing Curated Graph-Based Lexicographical Data. In J. Čibej et al. (eds.) *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts, 17-21 July, Ljubljana.* Ljubljana: University Press, pp. 1017-1022.

Meyer, P. (2019). Leistungsfähige und einfache Suchen in lexikografischen Datennetzen. Ein Query Bilder für lexikografische Property-Graphen. In P. Sahle (eds.): *Digital Humanities: multimedial & multimodal. 6. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019), Frankfurt am Main, Mainz, 25.3.2019 – 29.3.2019. Konferenzabstracts.* Frankfurt a.M., pp. 312-314.

Muysken, P. (2000). *Bilingual speech. A typology of code-mixing.* Cambridge: Cambridge University Press.

Newerkla, S. (2011). *Sprachkontakte Deutsch-Tschechisch-Slowakisch.* Frankfurt a. M., Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.

Poplack, S., Dion, N. (2012). Myths and facts about loanword development. In *Language Variation and Change* 24(3), pp. 279-315.

Simões, A., Salgado, A., Costa, R., & Almeida, J.J. (2019). LeXmart: A Smart Tool for Lexicographers. In I. Kosem et al. (eds.) *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal.* Sintra, Portugal. Brno, pp. 453-466.

Stachowski, K. (2011). A note on Levenshtein distance versus human analysis. In *Studia linguistica Universitatis Iagellonicae Cracoviensis 128.* Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, pp. 155-160.

Stutchkoff, N. (1950). *Der Oytser Fun Der Yiddisher Sprakh.* Accessed at:

http://www.cs.uky.edu/~raphael/yiddish/searchOytser.cgi [15/04/2021]

Walsleben, A. (1997). *Romanische Lehnwörter in polnischen Texten des 17. Jahrhunderts*. München: Kubon & Sagner.

LWP = Lehnwortportal Deutsch, ed. Leibniz Institute for the German Language. Accessed at: http://lwp.ids-mannheim.de [15/04/2021].

SALDAS = Sprachwissenschaftliche Arbeiten zu Lehnwörtern aus dem Deutschen in anderen Sprachen, ed. Leibniz Institute for the German Language. Accessed at: http://lwp.ids-mannheim.de/saldas [15/04/2021]

SGP = Słownik gwar polskich (A – hepnąć), Sources and vol. I, ed. M. Karaś, J. Reichan; vols. II-V, ed. J. Reichan, S. Urbańczyk; vol. VI, ed. J. Okoniowa, J. Reichan; vols. VII-VIII, ed. J. Okoniowa, J. Reichan, B. Grabka, vol. IX, ed. B. Grabka, R. Kucharzyk, J. Okoniowa, J. Reichan. Issue 1, Wrocław-Warszawa-Kraków-Gdańsk; since issue 2, Łódź; vol. III, Wrocław-Warszawa-Kraków; since vol. IV – Kraków, 1977-2017.

USTAWA = Ustawa o mniejszościach narodowych i etnicznych oraz o języku regionalnym z dnia 6 stycznia 2005 r. Accessed at: http://mniejszosci.narodowe.mswia.gov.pl/mne/prawo/zapisy-z-konstytucji-r/6447,Ustawa-o-mniejszosciach-narodowych-i-etnicznych-oraz-o-jezyku-regionalnym.html [14/06/2021].

WDLP = Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache. Von den Anfängen des polnischen Schrifttums bis in die Mitte des 20. Jahrhunderts, ed. A. Vincenz and G. Hentschel. Accessed at: http://diglib.bis.uni-oldenburg.de/bis-verlag/wdlp [15/04/2021].

WDLT = Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen, ed. Th. Menzel and G. Hentschel, second, amended and enlarged electronic edition 2005. Accessed at: https://www.bkge.de/Publikationen/Online/Woerterbuecher/Deutsche_Lehnwoerter_im_Teschener_Dialekt [07/04/2021].

Acknowledgements