Linguistica Silesiana 39, 2018
ISSN 0208-4228

MAREK KUŹNIAK
University of Wrocław
marek.kuzniak@uwr.edu.pl

HINDI, POLISH AND TRANSLATION. FACES OF LANGUAGE CONTACT IN THE CONTEMPORARY ENGLISH

This paper has two parts to it. The first part is about the presence and possible impact of Hindi and Polish as foreign words in the contemporary English language. This is measured via the proposed tool of CRAC (Cumulative Average Relative Count). The research is done on the basis of the British National Corpus (2001, 2007) and Longman Pronunciation Dictionary (2004, 2009). The focus is laid on the overriding heuristic metaphor LANGUAGE LAWS are PHYSICAL LAWS, where laws of lexical assimilation are viewed as analogous to physical laws of gravity. The second part marks the transition from a theoretical-descriptive perspective into a more practical, intercultural dimension. It is about translation of foreign proper names from the viewpoint of legal (certified) translation. This is a significant issue as many foreign words are actually proper names in English. This part relates then to specific controversies and proposed solutions concerning translation of Polish and Hindi proper foreign names in view of the presence and absence of their diacritic forms in English. The framework for adoption of the argument are institutionally established standards of certified translation practice in Poland.

Keywords: *language contact*, *metaphor*, *cognitive linguistics*, *translation*

1. Introduction – overview of Parts 1 and 2

This paper consists of two parts, theoretical-descriptive and practical. From the methodological point of view it combines the tools of cognitive linguistics and corpus-assisted analysis. The goal is both qualitative and quantitative and involves the reflection upon the contemporary status of Polish and Hindi as donor languages in regard of English as a recipient language. Qualitatively, the discussion hinges upon the tool of conceptual metaphor as a heuristic device to explicate the phenomenon of foreign lexis in the target language system. Quantitatively, the exploration draws on

comparison of some basic statistical data retrieved from the British National Corpus (BNC 2001, 2007) and lexical input obtained from the Longman Pronunciation Dictionary (LPD 2004, 2009). The data discussed in the first part are relevant especially to the discussion of foreign Polish and Hindi proper names from the viewpoint of certain requirements set by Polish recognized institutions for sworn translators in regard of the practices pertaining to the treatment of proper words.

Part 1

1.1. Gravitational face of language contact

In order to avoid possible interpretative ambiguities, I will use the term "foreign word" as introduced in Kuźniak (2009: 13) where the following prerequisites are contained:

> The word foreign is understood relative to a phonological criterion. Namely, if the word or phrase is uttered with the imitation of the pronunciation of the language of origin, this word or phrase counts as foreign. The information about such conceived 'foreignness' is brought by Longman Pronunciation Dictionary. The treatment of Longman Pronunciation Dictionary as the reference base for the elicitation of foreign words and phrases allows us to encompass, besides common words and phrases, a large group of proper foreign words and phrases for the analysis.

More specifically, the pre-selection of research data is done by encompassing a set of entries with a two-fold phonetic representation as shown in LPD (2004, 2009) below:

Andhra Pradesh
Andhra Pradesh <c mediumblue>ˌændr ə prɑː ˈdeʃ</c> ˌɑːndr-, -ˈdeɪʃ
<c green> —Hindi</c> \<<c mediumblue>aːn̪d̪ʱr prə d̪eːʃ</c>\
kielbasa
kielbasa <c mediumblue>kiː^əl ˈbɑːs ə</c> kɪl-, -ˈbæs-
<c green> —Polish</c> kiełbasa \<<c

In this paper I compare the results of the research discussed in Kuźniak (2009) based on LPD (2004) and the BNC World (2001) with the results drawn on the basis of LPD (2009) and the enhanced and upgraded the BNC XML (2007) edition. The BNC data[1] serve to demonstrate the frequency of occurrence of the pre-selected foreign words and phrases in the contemporary English. Certainly the data are treated with all necessary reservation, which is pointed out by Kuźniak (2009: 81):

[1] Data cited herein have been extracted from the British National Corpus, distributed by the University of Oxford on behalf of the BNC Consortium. All rights in the texts cited are reserved.

> However successful the BNC project may be, it is important to remark some significant caveats that cast an important light upon the results of research presented in the present book. The first reservation concerns the indiscriminate reliance upon the statistics about frequency provided by the BNC (see Aston and Burnard, 1998: 36–37). However, at no point throughout this book have we suggested that the statistics serve as unfailing evidence for drawing far-reaching conclusions. Rather the opposite has been implied. Thus, although the information about the frequency of occurrence of particular foreign words has led us to formulate some conclusive generalizations (Chapter 8, 9), these have, however, been made with the necessary proviso in view, i.e. the necessary idealization (simplification) of the presented results. This "idealization" prerequisite emerges from the awareness of various caveats hidden in the BNC. (…) The very "all-inclusiveness of the BNC" does not preclude other errors, especially the ones connected with the aforementioned problem of disambiguation. As Aston and Burnard (1998: 38) point out: "Quotations in languages other than English are also occasionally to be found, which may lead to confusion where they include forms which are identical to English words – for example, a fragment in German may contain many occurrences of the word die but have nothing to do with mortality".

The discussion in Kuźniak (2009: 20) is 'clothed' with the overarching heuristic metaphor LANGUAGE LAWS are PHYSICAL LAWS, thus providing the framework for the entire argument[2]. The argument goes like this:

> Language undergoes continual change. Change can be discussed in terms of motion. Language can thus be described as being in constant motion. Motion is determined by forces. Forces that determine language change are analogical to physical forces. There are two fundamental types of forces: centripetal (centre-seeking) and centrifugal (away-from-the-centre) forces. These physical forces are present in circular motion. Therefore language change can analogically be modelled as a circular motion. Languages being in constant motion can be compared to planets. Their mutual interactions are guaranteed by the centripetal force of interplanetary gravity. The atmosphere of a planet (e.g. the Earth [i.e. the English language]) is, however, the residue of opposite centrifugal-like forces. Once an alien body (e.g. a meteor), driven by the force of gravity, enters the Earth's atmosphere, it meets the resistance of densely accumulated air-particles that form the opposite (centrifugal-like) force(s) acting on this alien entity.

Obviously, I will not elaborate on these issues here as these are laid out in detail in Kuźniak (2009). Still, however, it must be noted that reconceptualising linguistic phenomena in terms of physical attributes is not a new endeavour. As early as in the 19th century with the advent of linguistics as a modern scientific discipline, attempts were made to equalize the laws of linguistic change

[2] The role of metaphorical mapping in everyday and academic thinking is best explained in Lakoff and Johnson (1980) and Lakoff and Johnson (1999).

with the laws of physics, especially at the level of phonology. On the plane of lexical-comparative studies it was Paul (1886), and the subsequent works by Seiler (1907-13) which also put the case for the overall spirit of the time (see Winford 2005: 42). The search for unification of the metalanguage of scientific description found its crowning point with the early 20th century neopositivist reductionist claim for physics as a point of conceptual-notional departure for other disciplines. Certainly, linguistic semiotics was also naturally affected by the dominant scholarly approach: As Wildgen (2008 [2000]: 75-76) argues:

> Many elements in Peirce's thought (for instance the 'marriage' between natural science and semiotics) and the importance of modern physics and modern experimental psychology for theoretical work in linguistics and semiotics, are also characteristic of Gestalt theory, the next stage in our journey to field-semantics.

Contemporarily, physics finds its adequate place as the feeding ground within the cognitive-linguistic approach. Worth noting here are the systematic studies by Talmy (2002) with his leading concept of Force Dynamics applicable to the discussion of the semantics of verbs (see also Kuźniak and Woźny 2014). Of significance are studies on physical substratum regarding the pre-conceptual image schemata found in Johnson (1987), Krzeszowski (1997), or Libura (2000). Interesting in this regard is Krzeszowski's (1997) extension of the discussion upon schemas to the area of valuation (cf. Kuźniak 2010). Also, Devitt and Sterelny in their *Language and reality. An Introduction to the Philosophy of Language* (1999: 10) add to the argument by putting matters rather straightforwardly: "Our theory of language must, therefore, be physicalist. Any linguistic facts there are must be, ultimately, physical. Semantic notions like meaning, truth, and reference can be used only if they can be explained in non-linguistic terms; they are not primitive". The inclination of contemporary linguistic theory towards the physical underpinning of natural language is best summed up by Krzeszowski (2012: 63) and his postulate for the explicative unifying metaphor that translates MENTAL REALITY in terms of PHYSICAL REALITY.

Putting aside the foregoing brief excursion into the history of contact between linguistics and physics, it should be sufficient, for the sake of the present argument, to quote the following basic correspondences, where physical force may be understood both as CENTRIPETAL (CENTRIFUGAL) FORCE OF GEOGRAPHICAL PROXIMITY (DISTANCE) between languages, and as CENTRIPETAL (CENTRIFUGAL) FORCE OF CULTURAL PROXIMITY (DISTANCE) between languages (Kuźniak 2009: 185). Because language is commonly agreed, since de Saussure at least, to be a social phenomenon, the discussed overarching metaphor is closely bound up with another conceptualisation where SOCIAL LAWS ARE PHYSICAL LAWS. Specifically, physical centrifugal and centripetal force may be compared to SOCIAL FORCE OF REJECTION or ACCEPTANCE, respectively. Along the same lines, physical force may be conceived as INSTITUTIONAL FORCE (media, academies, authorities) regulating the inflow of alien lexemes

into the target system, etc. It may thus be both centripetal or centrifugal, relative to circumstances (Kuźniak 2009: 185-186).

Having all this in mind, the following ontological/epistemic pairings can be further extracted (see Figure 1 below).

CONCEPTUAL WORLD
PHYSICAL WORLD ←————→ LINGUISTIC WORLD

PART I

Ontological correspondences	
The Earth	The English language
Planets	Major donor languages
Planetoids	Minor donor languages
Meteor	Lexeme
Mass of a meteor	CRACn1
Mass of a planet/planetoid	CRACn2

Figure 1. Ontological correspondences between physical and linguistic worlds (after Kuźniak 2009: 187)

Certainly, the key quantitative tools, i.e. CRACn1 and CRACn2 require explanation. These are acronyms from Cumulative Relative Average Count. As argued in Kuźniak (2009: 159):

> CRAC is, thus, "cumulative" because its value is always viewed as average of the sum of two parameters; it is "relative" because its value is always calculated relative to some other dimension; it is "average" because the mathematical calculation of "averaging" appears to be the closest conceptually to the cognitive process of blending (…) Finally; it is referred to as "count" because its value is the result of a mathematical calculation of "averaging".

Thus *mass of the individual meteor* (M) is calculated on the basis of averaging the number of hits in the entire corpus with the number of texts in which it is recorded. This yields CRACn1. For example, IF a given word occurs 5 times in the BNC in 3 texts, CRACn1 value is 4. CRACn2, on the other hand, is another technical term, which is significant for calculating the mass of the entire group of foreign words (proper or common) pertinent to a given language. Thus, assuming that after the pre-selection of foreignisms from LPD we have, say, 3 foreign proper Polish words in English where the word A has CRACn 1 equalling 5, the word B boasts CRACn1 equalling 7, and the word C has CRACn1 equal-

ling 260, then CRACn2 value is obtained as average of CRACn1 values of a given group. In the quoted hypothetical case, CRACn2 for Polish foreign proper words in English amounts to approximately 91. Then CRACn2, here 91, is multiplied by 3 (the hypothetical number of the Polish proper names in English), to yield the total value of M, which is 273. This procedure of CRAC elicitation is applied discretely to foreign proper words and common words.

Of relevance to this study is also the notion of *volume of the planet* (V), where it simply translates into total number of lemmas (proper or common) associated with a given language. So if, say, Polish has 20 foreign proper names according to LPD, V figure for this group equals 20. V and M figures are very significant for the entire argument, as V represents perceptually relevant criterion of 'size' of a planet, whereas M represents a planet's actual assimilatory potential in the target language landscape. Naturally, the two variables are independent of each other, and it may happen that the correlation between the two is either positive or negative. This, in turn, has significance for a tentative prognosis of the impact of a donor language on a recipient system. I will discuss that point at more length in Section 3 below.

1.2. Hindi and Polish as donor 'planets'

We are at the point where results of the research conducted on the basis of LPD (2004) and BNC World (2000) published in Kuźniak (2009) are confronted with the research results on the basis of LPD (2009) and the BNC XML (2007). Sections 2.1–2.8 illustrate lexical input retrieved in LPD (2004, 2009) according to the criteria laid out in Section 1, which were subsequently tested for their occurrence in the BNC (2001, 2007). The discussion is confined solely to Hindi and Polish[3].

1.2.1. Hindi proper names in English (LPD 2004, BNC World)

Agni; Allahabad; Amritsar; Andhra Pradesh; Arjuna; Asoka; Bihar; Bhagwan; Brahmaputra; Buddha; Deccan; Gandhi; Ganesh; Granth; Gujarat; Gujarati; Gupta; Gwalior; Haryana; Himalaya; Indore; Jaipur; Jalalabad; Karma; Lucknow; Mahabharata; Maharashtra; Maharishi; Mahayana; Marathi; Meerut; Nehru; Pathan; Patna; Pradesh; Rabindranath; Ramayana; Shiva; Sikh; Sind; Sindhi; Siva ; Srinagar; Tabla; Taj Mahal; Urdu; Vedanta

1.2.2. Hindi proper names in English (LPD 2009, BNC XML)

Agni; Allahabad; Amritsar; Andhra Pradesh; Arjuna; ***Arunachal****; Asoka;* ***Bhojpuri****;* ***Bhopal****;* ***Bihar****; Brahmaputra; Buddha;* ***Chindit****; Deccan; Gandhi; Ganesh; Granth; Gujarat; Gujarati; Gupta; Gwalior; Haryana; Himalaya;*

[3] Words in the bold type are new entries recorded in LPD 2009 / BNC XML editions in comparison to the previous LPD 2004/BNC World versions.

Indore; Jaipur; Jalalabad; ***Khalsa****; Lucknow; Mahabharata; Maharashtra; Maharishi; Mahayana; Marathi; Meerut; Nehru; Pathan; Patna; Pradesh; Rabindranath;* ***Rajpoot****; Ramayana; Shiva; Sikh;; Sind; Sindhi; Siva; Srinagar; Taj Mahal; Urdu;* ***Uttar Pradesh****; Vedanta.*

1.2.3. Hindi common words in English (LPD 2004, BNC World)

babu; basmati; bhangra; bhindi; chapatti; chela; dak; deva; dhal; dhansak; dharma; dhobi; dhoti; garam masala; guru; hanuman; hatha; jai; lakh; lassi; maharaja; maharanee; mahatma; nirvana; pandit; puggree; puja; purda; raj; roti; sandhi; saree; sari; satyagraha; sitar; suttee; svarabhakti; swami

1.2.4. Hindi common words in English (LPD 2009, BNC XML)

Babu; basmati; ***bhagwan****;* ***bhaji****;* ***bhang****; bhangra;* ***bharal****; bhindi;* ***bindi****; chapati; chela; dak; deva; dhal; dhansak; dharma; dhobi; dhoti; garam masala; guru; hanuman; hatha; jai;* ***kabaddi****;* ***karma****; lakh; lassi; lathi; maharaja; maharanee; mahatma; nirvana; pandit; puggree; puja; purda; raj; roti; sandhi; saree; sari; satyagraha; sitar; suttee; svarabhakti; swami;* ***tabla***

1.2.5. Polish proper names in English (LPD 2004, BNC World)

Bialystok; Bydgoszcz; Cracow; Gdansk; Gorecki; Jan; Jaruzelski; Katowice; Kosciusko; Krakow; Lodz; Paderewski; Pilsudski; Rzeszów; Strzelecki; Szczecin; Torun; Vistula; Wajda; Walesa; Wojtyla; Wroclaw

1.2.6. Polish proper names in English (LPD 2009, BNC XML)

Bialystok; Bydgoszcz; Gdansk; Gorecki; Jan; Jaruzelski; Katowice; Kosciusko; Krakow; Lodz; Paderewski; ***Penderecki****; Pilsudski;* ***Poznan****; Rzeszów; Strzelecki; Szczecin; Torun; Vistula; Wajda; Walesa; Wojtyla;* ***Wozniak****; Wroclaw*

1.2.7. Polish common words in English (LPD 2004, BNC World)

grosz; kielbasa; mazurka; zloty

1.2.8. Polish common words in English (LPD 2009, BNC XML)

grosz; kielbasa; mazurka; ***ogonek****; zloty*

1.3. Hindi vs. Polish

In this part I compare Hindi and Polish as donor planets relative to the recipient English language. The aforementioned categories of V and M are implemented for their qualitative-quantitative heuristic value, where, as stated above, V is the total of lemmas recorded in a donor language. V is subdivided into V1 and V2, i.e. the proper names and common words, respectively. M, in turn, is obtained from the aforementioned CRAC operations. M, analogically to V, is respectively subdivided into M1 (proper names) and M2 (common nouns)[4].

Table 1.
(adapted from Kuźniak 2009: 291 [based on LPD 2004 and BNC World 2001])

Language	V1	V2	V	M1	M2	M
Hindi	47	38	**85**	1598	608	**2206**
Polish	20	4	**24**	1260	40	**1300**

1.3.1. Hindi, Polish and other 'planets'

In order to see how Hindi and Polish 'fill in' the foreign lexical landscape of English it is advisable to introduce more representative statistics including the most influential 'planets'. Not surprisingly, French leads the ground with the greatest M value. Interesting is the case of Arabic, especially in comparison to Welsh. M value of Arabic is much higher than that of Welsh. A tentative conclusion is that prospects for assimilation of Arabic words in English are much greater than in the case of Welsh. The assimilatory force, which is associated with M, emerges somewhat naturally as a corollary of physicalist nature of language steered by gravity laws. In other words, the greater the mass of a donor planet, the greater its assimilatory potential towards the recipient system. This is 'guaranteed' by Newtonian physics. As can be seen from Table 2 and 3 below, M of Hindi is greater than M of Polish. We may thus conclude that Hindi bespeaks its higher global placement when it comes to contact with English.

[4] As Kuźniak argues (2009: 289): "the entire mass of a planet was linguistically correlated with the following formula: M1= (Ln x CRACn2 [y]); M2= (Ln x CRACn2 [y]); M= M1+M2, where *M1* relates to the value obtained on the basis of the formula that reads: the number of proper names or phrases (Ln) times (x) CRACn2 value (y) calculated for the relevant set. It is to be reminded that CRACn2 value is the averaged set value calculated on the basis of CRACn1 values of a set of lemmas pertinent to a particular category, whether 'proper names' or 'common words'. *M2*, on the other hand, relates to the value obtained on the basis of the formula that reads: the number of common words or phrases (Ln) times (x) CRACn2 value (y) for the relevant set. *M* stands for the sum total of *M1* and *M2*".

Table 2. Hindi, Polish and other 'planets'
(adapted from Kuźniak, 2009: 290–291)

Language	V1	V2	V	M1	M2	M
French	547	598	**1146**	28991	19734	**48725**
German	414	91	**505**	25668	1547	**27215**
Italian	210	109	**319**	9450	1853	**11303**
Spanish	215	65	**280**	16125	1690	**17815**
Welsh	224	7	**231**	4704	63	**4767**
Russian	148	15	**163**	8140	420	**8560**
Chinese	109	8	**117**	4469	352	**4821**
Hindi	47	38	**85**	1598	608	**2206**
Dutch	63	0	**63**	4410	0	**4410**
Japanese	35	28	**63**	2870	532	**3402**
Arabic	42	6	**48**	7854	456	8310
Portuguese	39	5	**44**	1248	30	**1278**
Danish	21	3	**24**	609	42	**651**
Greek	21	11	**32**	357	165	**522**
Swedish	21	4	**25**	1050	12	**1062**
Polish	20	4	**24**	1260	40	**1300**

Table 3. Hindi vs. Polish
(present research [LPD 2009, BNC XML 2009 edition]

Language	V1	V2	V	M1	M2	M
Hindi	51	47	**98**	1683	676	**2359**
Polish	24	5	**29**	1320	42	**1362**

As noted above, the physicalist cognitive linguistics, whose assumptions were preliminarily presented in Kuźniak (2009, 2013) proposes that language, as any other physical object in the world, is subject to gravity laws. This further implies a number of ontological-epistemic correspondences, some of which were already laid out in Section 1. When the results of lexical contribution of Hindi and Polish from Tables 1-3 are compared, we predictably do not observe any vehement changes as

to the actual impact the languages in question have on English. The increase of M value in the case of Polish amounts to 0,045%, whereas in the case of Hindi the increase is slightly greater and oscillates round 0,065%. Strangely, when it comes to V, Polish records a double increase compared to Hindi. This, however, has so far a marginal significance as to the assimilatory potential of Polish lexis in English, as M indicator unfailingly demonstrates. The difference in interpretation of data arises if we take into account the working notion of Impact Factor (IF), which marks the prognosis for the prospective degree of impact of a donor language with respect to English as a recipient language (see Table 4 below). IF is calculated as a quotient of M divided by V. This is further discussed in Section 3.2.

Table 4. Most influential 'planets'. Potential IF of Polish and Hindi on English in BNC World 2001 and BNC XML 2009 edition (in square brackets)

Planet	**IF (M/V)**
Spanish	**63,6**
German	53,9
Polish	54,1 [46,9]
French	42,5
Italian	35,4
Hindi	24,1 [25,9]
Welsh	20,6

1.3.2. The problematic Jan

The results shown above allow for the methodology applied in Kuźniak (2009), where it was the first language identification of a given multilingual foreign word in LPD which counted as a definitive assignment relevant to the research. The rationale behind the method was the materialist, perceptual criterion for the elicitation of foreign lexis based on the aforementioned, mostly twofold pattern of pronunciation of a given entry in English. Consistently then, a proper noun *Jan* was assigned the Polish origin and classified as such in the study. To illustrate that point, LPD original annotation is again provided:

Jan
Jan <c mediumblue>dʒæn</c> —but as a male name also <c mediumblue>jæn</c> jɑːn —and by confusion also ʒð, ʒæn<c green> —Polish, Czech, Swedish, Norwegian</c> \<<c mediumblue>jan</c>\><c green>, Dutch</c> \<<c mediumblue >jɑn</c>\>

As the entry illustrates, *Jan*, although primarily designated as Polish foreign name, is also ascribed to Czech, Swedish, Norwegian and Dutch. This has its consequences for Polish M and the consequent IF calculations. As Table 4 demonstrates, non-disambiguated *Jan* contribution to IF of Polish is roughly twice as big as the overall IF of Hindi. If we, however, disambiguate *Jan*, and explore, which is not that certain in many cases, the Polish context of the word occurrence, we receive the following general statistics:

- disambiguated *Jan*, hits: 71, texts: 31, CRACn1: 51
- non-disambiguated *Jan*, hits: 1415, texts: 406, CRACn1: 910,5.

The consequences for Polish as the donor language are quite telling and inarguably show up the weaknesses of the methodology, which adopts a unanimous approach to all entries by neglecting disambiguation effects. As it was said above, such procedure is justified in view of the perceptual (non-semantic) criterion of elicitation of data as well as controversies with disambiguation procedure as such. Still, however, disambiguation plays a big role in interpreting comparative data. On adopting a non-disambiguation procedure, Polish has a greater IF than Hindi, while, on the application of disambiguation mechanism, IF of Polish is lower than IF of Hindi and amounts to 17,1, with M equalling 485, and V equalling 29 (LPD 2009, BNC XML). Interestingly, if the disambiguation procedure for *Jan* is applied to the original data drawn on LPD (2004) and the BNC World (2001), the revised statistics for Polish would present the following: IF: 18,6, M: 446; V: 24. An optimistic conclusion is that whatever the methodology applied (disambiguation or non-disambiguation), the overall tendencies of the impact of Polish on English remain roughly the same: IF of Polish (LPD 2009, BNC XML) is slightly lower than IF of Polish (LPD 2004, BNC World) with M about 8% higher (LPD 2009, BNC XML) than M obtained from LPD 2004 and the BNC World data. The data may be distorted in conducting a comparative cross-linguistic analysis if we neglect disambiguation effects, but the same effects have no relative significance for measuring M/V intra-linguistically. Without taking into account disambiguation effects, the revised IF of Polish (LPD2009, BNC XML) is about 12-13% lower than IF of Polish (LPD 2004, BNC World). Disambiguation effects applied, the revised IF of Polish (LPD2009, BNC XML) is about 8-9% lower than the IF of Polish (LPD 2004, BNC World). All in all, the observed problems apply to Polish only and have not been detected for other donor languages, which encourages the assumption that the overall methodology is by and large operative.

Statistics presented in Tables 1-4 appear as somewhat predictable results, given the naturalist entailments uncovered throughout the study (Kuźniak 2009), where laws of lexical assimilation and laws of gravity showed a great deal of correspondences, even at a fine-grained corpus-assisted level of description. It follows then that if planets evolve rather slowly, imperceptibly to humans, this is also the case with language, or more precisely its lexical system. It should be emphasised that methodological problems of disambiguation which were revealed in the case of *Jan* may be statistically relevant to the interpretation of the

'size' of Polish in relation to English, which is half as big as originally projected. Still, however, the new 'corrected' statistics actually confirm the tendencies observed for Polish, no matter whether or not the revised methodology taking disambiguation effects for granted is applied.

Part 2

2. Legal faces of language contact

On the web page of the Polish Society of Sworn and Specialized Translators we read[5]:

> As a result of the cooperation of TEPIS and the Ministry of Justice, 'The Sworn Translator's Code' was prepared in 2005, then verified jointly with representatives of other ministries and universities in 2011. That was a great success for the profession and for the implementation of recommendations made by AGIS, a project of the European Commission. The contents of the Code include rules of ethics and good practice for this profession. It is also a kind of implementation of the 2004 Act on the Profession of the Sworn Translator.

From the perspective of research results presented in Sections 1-3, of interest to us is the actual status of Polish and Hindi words in English. Certainly, the aforementioned Code is a highly influential, albeit not binding, document for the community of sworn translation professionals in EU. Still, however, the problems that arise in translation practice appear to be of more universal nature.

In § 37 of The Code, the following information concerning diacritic marks in proper names can be found:

(1) Proper names containing diacritic marks in the source language preserve these marks in the target language translation.
(2) Failure to contain alleged diacritic marks by proper names in the translation from the source language shall be noted by the translator.
(3) Personal proper names and geographical names which contain no diacritic marks in the source language shall be consistently quoted without diacritics in the target language system.

Let us now confront the postulates enshrined in the Code with the practices of sworn translators in Poland. At first sight the problem of translating proper names in English-Polish translation is unambiguously resolved by the Code. The translation practitioners are very much 'invited' to quote proper names as they appear in the source language. At the same time, however, the same practitioners are also 'invited' to draw on a vast repertoire of corpora in making

[5] http://www.tepis.org.pl/index.php/tepis-o-sobie/dzialalnosc-pt-tepis/the-polish-society-of-sworn-and-specialized-translators.

translation choices in search of their empirical validity. One such expert corpus is definitely the BNC for the Polish translators.

When we look, however, at the list of foreign proper names, we notice that out of the following names: *Bialystok; Gdansk; Gorecki; Kosciusko; Krakow; Lodz; Pilsudski; Poznan; Rzeszów; Torun; Walesa; Wojtyla; Wozniak; Wroclaw*, the only one that preserves its diacritic marks in LPD (2004, 2009) is *Rzeszów*. Other names are left without diacritics. This is a very telling story. When we, in turn, insert the names into XAIRA search engine along with their full Polish graphical make-up, we receive the following statistics of occurrence (hits/texts, respectively): *Białystok* (0/0), *Gdańsk* (3/1), *Górecki* (0/0), *Kościuszko* (1/1), *Kraków* (4/2), *Łódź* (0/0), *Piłsudski* (11/1), *Poznań* (2/1), *Rzeszów* (0/0), *Toruń* (0/0), *Wałęsa* (0/0), *Wojtyła* (0/0), *Woźniak* (0/0), *Wrocław* (0/0).

Now, let us see the statistics for the same words spelled without diacritics: *Bialystok* (3/2), *Gdansk* (62/33), *Gorecki* (1/1), *Kosciusko* (1/1), *Krakow* (49/31), *Lodz* (18/10), *Pilsudski* (4/3), *Poznan* (23/16), *Rzeszow* (2/2), *Torun* (2/1), *Walesa* (264/60), *Wojtyla* (6/6), *Wozniak* (3/3), *Wroclaw* (14/7). As illustrated in Table 3 above the contribution of Polish proper names to M1 is 1320. If we include Polish proper names that occur in their diacritics in the BNC corpus, the contribution increases to 1344, which constitutes only 2% of such extended corpus. We may, therefore, conclude that the role of Polish proper names marked with diacritics is marginal if not statistically negligible. This certainly clashes with the rather strong statements from TEPIS about the necessity to indiscriminately translate names from Polish into English along with their original graphic representation containing diacritics. The story becomes even more sensitive when you translate *Wałęsa* or *Wojtyła* into English. Both persons are known worldwide. The former, the legendary Solidarity anti-communist leader, the latter, Pope John Paul II. There is every reason to suppose that these names should preserve their original spelling in English when translated from the source language. The BNC statistics leave no doubt that the opposite is true. What is more, the fact that these names lose diacritic marks in English texts is consonant with the physicalist view expounded in Kuźniak (2009), where the processes of assimilation with the end-point of adaptation into the target lexical system is not cost-free and normally results in the alterations in the graphemic composition of the word and may, depending on the complex usage circumstances, fossilize as such in the target system. Alternatively, a native-like equivalent is coined to satisfy communicative needs of the target system community (e.g. Warsaw).

When we take a look at Hindi in contact with English, diacritized forms may appear in proper names provided they are ancient names, not modern ones. Hence we receive: *Mathurā, Kauśambī, Valabhī, Kāñcī*, *Uraiyūr, Tiḷevalli* etc., but *Allahabad* (not *Allāhābād*), *Kolkata/Calc*utta (not *Calcaṭṭā*), *Chennai/ Madras* (and not *Madrāsa*).[6]

[6] Guide to transliteration: www.springer.com/cda/.../Guide+to+Transliteration.pdf?

Table 5. Hindi non-diacritized proper names with the BNC statistics of occurrence and their diacritized equivalents

LPD form	Hits	texts	Transliterated form	Hits	Texts
Amritsar	25	15	Rāmdāspur	0	0
Andhra Pradesh	38	19	Āṁdhra Pradeś	0	0
Arunachal	15	7	Aruṇācal Pradeś	0	0
Asoka	4	4	Aśoka	0	0
Bhojpuri	2	2	Bhodźpuri	0	0
Bhopal	50	25	Bhopāl	0	0
Bihar	80	31	Bihār	0	0
Gandhi	677	113	Gaṁdhī	0	0
Gujarat	40	22	Gujarāt	0	0
Gwalior	9	3	Gvāliyar	0	0
Haryana	51	17	Hariyāṇā	0	0
Khalsa	3	3	Xālisa	0	0
Lucknow	20	17	Lakhnaū	0	0
Mahabharata	19	10	Mahābhārata	0	0
Mahayana	9	8	Mahāyāna	0	0
Marathi	2	2	Marāṭhī	0	0
Patna	16	9	Paṭnā	0	0
Pradesh	199	55	Pradeś	0	0
Ramayana	3	3	Rāmāyana	0	0
Urdu	55	33	Urdū	0	0
Uttar Pradesh	95	41	Uttar Pradeś	0	0
Vedanta	0	0	Vedānta	1	1

A cursory glance at Table 5 above is sufficient for us to formulate rather strong statements about the lack of representation of diacritized equivalents of Hindi proper names in the BNC, a noble exception being *Vedānta*. The problem of diacritized forms in Hindi is different than in Polish on account of the accepted threefold realisation of spelling: on the one hand, we have Devanagari

alphabet in which Hindi is originally written, while, on the other hand, we have potential transliterated versions written in Roman alphabet and their English non-diacritized equivalents. Certainly, we have to bear in mind that English enjoys the status of an accessory language of communication in India. This is probably one of the reasons why the problem of translating proper names into English is practically non-existent, and if any tips for the translator may appear useful, these rather concern a clear corpus preference for Hindi non-diacritized proper names in the BNC.

2.1. The case of Lech Walesa International Airport

Closing up the discussion about translation of proper names, let us examine one interesting Polish case with rendering the official name of the Gdańsk Airport in English, as this issue was raised by the Polish Language Council – an institution appointed by force of the Act on the Polish Language from 1999. According to Article 10, par. 1 of the Polish Language Act of October 7th 1999, official names of state infrastructure shall be given priority over other renderings (see Kuźniak and Mańczak-Wohlfeld, 2016).The problem with the airport name was that it gave priority to the English version. Additionally, in the course of a heated debate over the issue, the problem cropped up whether *Wałęsa* shall be rendered as Walesa in English or in original diacritized forms. The argument put forward by the Council was that Lech Walesa may refer to a different individual than *Lech Wałęsa*, so the insistence was on the provision of diacritized form in the official airport name. In view of what we said above, the Council acted in consistence with the recommendations of the specialist translators' communities (the aforequoted TEPIS), with the strong implication that the Polish name of the airport, i.e. *Port Lotniczy Gdańsk im. Lecha Wałęsy* does not only meet the spelling, and grammatical requirements of the Polish Language, but also unequivocally designates the right individual referred to by the diacritized form *Lech Wałęsa*. The management of the *Gdańsk Airport* did not, strangely enough, respond to the official letter from the Council, enforcing the aforementioned change.

The whole debate appears eventually not to have been resolved, but in view of the corpus-based approach to translation the discussion over the identity disambiguation connected with *Lech Walesa* and *Lech Wałęsa* seems at least arguable, at most flatly pointless. Consistently then, in accordance with the corpus-based research findings, the Polish form *Lech Wałęsa* should be rendered *Lech Walesa* in English, whereas in translation from English into Polish *Lech Walesa* should be translated as *Lech Wałęsa*. This is not in line with the recommendations by the authoritative bodies in Poland, but the rule, as any commonsense-based strategy, may be applied flexibly and vary with the proper names depending on the referent, the final decision on a diacritized or non-diacritized version is left to the translator, especially in certified translation, where the person involved in the business is by force of statute a person of public confidence.

This rule of translating proper names may be extended to documents, where preservation of diacritic/non-diacritic forms in the target language may be obligatory, as a rule, to such instruments as vital record certificates (birth/death/marriage/school certificates, court judgements, and other types of instruments, where failure to quote a proper name in the target language may produce certain legal effects. So if the name *Wozniak* appears on the dissolution decree, it should be quoted as *Wozniak* in the Polish version. Conversely, if *Woźniak* appears on a Polish dissolution decree, it should indiscriminately be rendered as *Woźniak* in the English version. It does not mean, however, that certain proper names (geographical names, city names, or personal names) which are not the integral part of the document shall be treated with the same rigor. The assessment of whether a given proper name is, or is not, an integral part of the document is finally left to the translator. It is hard to imagine, for example, leaving in English-Polish translation the non-diacritized *Wroclaw* as the Polish city name in the target Polish documents, where *Wrocław* formally remains in accordance with conventions of the Polish spelling system. On the other hand, it would be received as hypercorrect if not snobbish to translate into English *Wrocław* as *Wrocław* in business correspondence, where *Wroclaw* is the element of, say, promotion campaign of the city in international relations[7].

2.2. Green light in the tunnel

The update of the 2004 Act on the Profession of Sworn Translator from 2015 (Article 14[2]) comes as a partial solution to the problems of translating diacritized proper names. Statutory provisions stipulate that the sworn translator may request the verification of the spelling of a name or surname on the basis of some other document (passport, ID) in the case when the doubt arises how to render that name in translation. The issue particularly concerns translation from the source languages operating in a non-Latin alphabet. However, the wording of the quoted Section has been extended in practice to cover the cases of English-Polish translations, particularly in documents, where accuracy in rendering personal data is absolutely crucial for the validity of the document (e.g. birth or death certificates, marriage certificates, certificates of no impediment, UK; US driving licences, tax returns [UK P60 forms], car registration documents). On a very practical, everyday basis Polish offices usually accept the additions of Polish diacritics to the non-diacritized names in the source language documents, provided the scan or copy of the document verifying personal data is enclosed with the translation. This flexible 'policy' is generally observed in large cities. The local authorities operating in smaller communes, however, remain rather inelastic when it comes to 'improving' personal names in translation. They often put down their inflexibility to the 'fear' of external audit form higher

[7] See the official English website of the city of Wroclaw (http://www.wroclaw.pl/en) with the consistent use of non-diacritized form of the city name.

administrative bodies. Last but not least, the translator, who faces the dilemma of whether to convey the source language form with diacritics or not, may be confronted with the clients who may refuse to provide additional verification documents, claiming that they do not have the obligation to do so. The rationale behind such refusals, rare as they are, is that the sworn translator's task is to offer professional services, which means they know how to proceed in a difficult case without causing any disturbance to clients, who may wonder whether the translator does not exceed his/her powers asking them for additional identification documents.

3. Conclusions

Foreign lexis in English behaves like a constellation of planets with the interplanetary gravitational force regulating the degree of impact of planets with respect to the biggest planet "the Earth", here English. From that point of view, Hindi looks like an average planet with Polish enjoying the status of at most a planetoid in relation to English. This is certainly an anthropocentric perspective conjoining facts of the Newtonian physics with Ptolemy's geocentrism. This combination of expert and naive perspectives is rather a rule than exception. Take, for example phrases like 'the sun rises' or 'the moon lights', where an anthropocentric view of the world is founded on a bodily-based entrenchment of the human conceptualiser in the environment exists parallel to the expert views expounded in formal institutionalised contexts. The two worlds are not mutually exclusive but rather complementary, forming a synergistic continuum of human cognition. The story about lexical contacts between Hindi, Polish and English appears to be strangely significant to the translation enterprise, especially in certified legal (court) translation of proper names in Poland. Foreignisms are a big dilemma in this respect since their status is not fully determined both at the level of phonological and graphic composition. This is specially manifested in the presence or absence of diacritics, which is usually the first indication of the adaptation process of a foreign word in the target language. The BNC shows that quite clearly. These corpus-based findings are in opposition to the recommendations by various authoritative institutions in Poland that require that translators quote in their entirety proper names in the target language. At the same time, the translations are encouraged at various professional upgrade courses to regularly verify their choices with the aid of corpus data. This produces some sort of cognitive dissonance, so it seems that these clashes can be resolved only locally with the support of a common sense rather than based on general, exceptionless rules.

References

Aston, G., and L. Burnard 1998. *The BNC handbook. Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Devitt, M., and K. Sterelny [1987] 1999. *Language and reality. An introduction to the philosophy of language*. Cambridge, MA: MIT Press.

Kierzkowska, D. (ed.) 2005. *Kodeks tłumacza przysięgłego z komentarzem* [The Sworn Translator's Code with commentary]. Warsaw: Translegis.

Krzeszowski, T.P. 1997. *Angels and devils in hell. Elements of axiology is semantics*. Warszawa: Wydawnictwo Energeia.

Krzeszowski, T.P. 2012. *Meaning and translation*. *Part 1: Meaning*. Frankfurt am Main: Peter Lang.

Kuźniak, M. 2009. *Foreign words and phrases in English. Metaphoric astrophysical concepts in lexicological study*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Kuźniak, M. 2010. On electromagnetic 'space' and linguistic valuation. In A. Ciuk and K. Molek-Kozakowska (eds.), *Exploring space: Spatial notions in cultural, literary and language studies; Volume 2. Space in language studies*, 62-74. Newcastle upon Tyne: Cambridge Scholars Publishing.

Kuźniak, M. 2013. Fizykalizm w semantyce kognitywnej. Trzecia droga? In P. Stalmaszczyk (ed.), *Metodologie językoznawstwa V*, 83-102. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Kuźniak, M., and J. Woźny 2014. Linguistic force dynamics and physics. In K. Rudnicka-Szozda and A. Szwedek (eds.), *Cognitive linguistics in the making*, 145-160. Frankfurt am Main: Peter Lang.

Kuźniak, M., and E. Mańczak-Wohlfeld 2016. Language planning activities and policy – the case of Poland. *Glottodidactica*. Vol. XLIII/2, 63-83. Poznań 2016: Wydawnictwo Naukowe UAM.

Lakoff, G., and M. Johnson 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.

Lakoff, G. 1993. The contemporary theory of metaphor. In A. Orthony (ed.), *Metaphor and thought*, 202-252. Cambridge: Cambridge University Press.

Lakoff, G., and M. Johnson 1999. *Philosophy in the flesh. The embodied mind and its challenge to western thought.* New York: Basic Books.

Libura, A. 2000. *Wyobraźnia w języku. Leksykalne korelaty schematów wyobrażeniowych CENTRUM-PERYFERIE i SIŁY*, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Paul, H. 1886. *Prinzipien der Sprachgeschichte*. Halle: Max Neimeyer.

Seiler, F. 1907-13. *Die Entwicklung der deutschen Kulture im Spiegel des Lehnwortes. 4 vols. Halle: Buchhandlung des Waisenhauses*.

Talmy, L. 2002. *Toward a cognitive semantics.* Vol. I. Cambridge, Massachusetts: The MIT Press.

Wells, J. (ed.). 2004 [1997]. *Longman pronunciation dictionary*. 2nd edition. London: Longman.
Wells, J. (ed.). 2009 [2007]. *Longman pronunciation dictionary*. 3rd edition. London: Longman.
Wildgen, W. 2000. The history and future of field semantics. From Giordano Bruno to dynamic semantics. In L. Albertazzi (ed.), *Meaning and cognition: A multidisciplinary approach*, 203-226. Amsterdam, Philadelphia: John Benjamins.
Winford, D. 2005. *An introduction to contact linguistics*. Oxford: Blackwell Publishing.

Electronic sources

Act of 25 November 2004 on the Profession of Sworn Translator (promulgated in Dziennik Ustaw 2004, No. 273, item 2702). Retrieved October, 31, 2017 from: http://www.mc-tlumaczenia.pl/files/file/PolishSwornTranslatorAct(27_1_2005).pdf
Guide to transliteration. Retrieved September, 29, 2016 from http://www.springer.com/cda/.../Guide+to+Transliteration.pdf?
Polish Society of Sworn and Specialized Translators. Retrieved October 20, 2017 from: http://www.tepis.org.pl/index.php/tepis-o-sobie/dzialalnosc-pt-tepis/the-polish-society-of-sworn-and-specialized-translators.
The British National Corpus, version 2 (BNC World). 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/
The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/