1 Introduction

It is a linguistic and anthropological truism that each natural human language carves out its unique lexical landscape. The resulting differences between languages constitute attractive articles of edutainment. It is indeed both amazing and amusing that the same dessert known as a *floating island* in English and in Spanish (albeit pluralized, as *islas flotantes*) is called "snow eggs" in German (*Schnee-Eier*), "nothing soup" in Polish (*zupa nic*), and "bird's milk" in Romanian (*lapte de pasăre*). In a situation somewhat resembling a projection test in psychometry, we are confronted with different projections of extralinguistic images in different languages. In some of them, such as French, two competing projections are found, viz. "eggs on the snow" (*œufs à la neige*) and "floating islands" (*îles flottantes*). In a similar example, the "at" sign (@) is a little monkey, snail, duckling, elephant's trunk, dog, etc. if we move from one language to the other (see http://europapont.blog.hu/2014/06/12/at_around_europe_bigger; accessed November 17, 2014).

Similarly fascinating is the fact that Mandarin Chinese does not construe *brother* as an integrated concept, insisting rather on separate words for "younger brother" (*dìdì,* 弟弟) and "older brother" (*gēgē,* 哥哥). Meanwhile, speakers of various Slavic languages do not have to differentiate between *leg* and *foot*, calling both *noga* (with *g* changing into *h* in some of them) – a feature that they share with the Bavarian dialect of German, in which *fuas* is used for both, unlike in standard German, in which one differentiates between *Bein*, "leg," and *Fuß*, "foot."

However fascinating these cross-linguistic lexical differences may be, linguistics, being a social and behavioral science, is not about entertaining people but, rather, about discovering generalized regularities and finding practical solutions to real-world problems. Consequently, the goal of this linguistic monograph is to go beyond fascinating and interesting, and attempt to find patterns and solutions in detecting and resolving linguistic conflict stemming from cross-linguistic lexical differences. Its first part, titled "Toward a Taxonomy of Cross-Linguistic Lexical Differences," which follows this introduction, represents an attempt to outline configurations of differences, taking various linguistic, cognitive, and cultural parameters into consideration. In contrast, Part II of the book, "Lexicographical Considerations," examines the

1

solutions to the challenges that cross-linguistic lexical differences bring about
in lexicography. The final chapter of the manuscript summarizes the findings of
the two parts and offers an outlook on the applications beyond lexicographic
strategies.

The present-day global dominance of the English language is another lin-
guistic and political truism. On the surface, English is the most commonly
taught language, the language most translated into and out of, and the language
with most bilingual dictionaries. At a deeper level, the linguistic dominance of
English means that numerous speakers of various other languages are forced to
use English and hence immerse themselves in the lexical and conceptual land-
scape of the English language, which may be substantially different from that
of their native language. The latter is abundantly documented by Wierzbicka
(2013), who points out important conceptual differences between English and
numerous other world languages, and argues for a transnational and transcul-
tural social science based on a universal metalanguage, rather than one domi-
nant language and its conceptual framework. While it remains to be seen
whether this noble endeavor will take any traction in real life or remain limited
in its influence, following the destiny of Esperanto, Ido, and other similar
proposals, Wierzbicka's depiction of the global dominance of the English
language is incontrovertible.

For all these reasons, the present monograph will focus mostly on the cross-
linguistic lexical differences between English and various other languages. To
develop this point further, highlighting the differences between English and
other languages is a contribution (however minute it may be) toward preserving
the linguistic and cultural diversity of our world.

The attempt to recognize and catalogize the aforementioned differences
follows Dumont (1970: 249): "The oneness of the human species, however,
does not demand the arbitrary reduction of diversity to unity—it only demands
that it should be possible to pass from one particularity to another, and that no
effort should be spared in order to elaborate a common language in which each
particularity can be adequately described. The first step to that end consists in
recognizing differences."

Determining the scope of analysis is a sine qua non of any scholarly analysis.
In this particular case, the scope is determined by the operational definition of
cross-linguistic lexical differences. Cross-linguistic lexical anisomorphism, to
use a more technical term for these differences, belongs to the class named
notational terms – that is, to terms that can "be defined differently in different
frameworks" (Lipka, 1992: 5).

What, then, is the framework of this monograph? Broadly speaking, CLA is
defined within the general concept of the "post-methodology era" (see Brown,
2002, for more information). The key premise of a "post-methodology"
approach is the utilization of any valid elements from any available approach

that can contribute to the overarching scholarly goal – in this case, finding the regularities in CLA and the solutions for the challenges that CLA causes. A similar approach is shared in some recent studies on the lexicon, in particular by Hanks (2013). Like the approach that Hanks advocates, the present study is also driven by empirical data and the principle that the flow of the analysis is bottom-up. All presented metadata are extracted from the analysis of comprehensive real-life data sets gathered in various relevant dictionaries, monographs, and other sources.

The nature of the problem makes contrastive linguistics the first theoretical framework that needs to be incorporated into this monograph. I adopt here the view advocated by Gast (2012: 140), namely that "contrastive analysis investigates the differences between pairs (or small sets) of languages against the background of similarities and with the purpose of providing input to applied disciplines such as foreign language teaching and translation studies." Another useful set of ideas comes from systemic functional linguistics (Webster, 2009: 5):

Language, like other semiotic systems, is a systemic resource for making and exchanging meaning. Language is a particular kind of semiotic system, which is based on grammar, characterized by both a stratal organization and functional diversity. Both this stratal organization and metafunctional diversity in language combine to form what M. A. K. Halliday refers to as a semiotic of higher-order consciousness, the basis for the human activity of meaning.

Language is the instantiation of an indefinitely large meaning potential through acts of meaning that simultaneously construe experience and enact social relationships. Acts of meaning are the linguistic instances of the linguistic system of meaning potential. Acts of meaning are a subclass of semiotic acts that are semantic.

A semantic system is a system of meaning, which is distinguished from other semiotic systems by the fact that it is founded on grammar. It is a system of meaning of a natural language, a system of wordings. The semantic system is one of three levels, or strata, which together comprise the whole linguistic system. Between the semantic system above and the phonological and morphological realization below is the lexicogrammar.

For an extended discussion about Halliday's work, see Kilpert (2003). Of particular importance here is the idea about the social embeddedness of linguistic meaning and the place of the lexicogrammar in the overall linguistic model.

Finally, my framework is informed by cognitive and cross-cultural linguistics, and in particular by the following propositions.

(1) The idea that "[t]he conventional meaning of a lexical item must be equated with the entire network, not with any single node" (Langacker, 1991: 3).
(2) The binary (phonological-semantic) model of symbolic units, as presented by Langacker (1991).

(3) The role of metaphor in language, as presented by Steen (2007), Kövecses (2005), and Lakoff and Johnson (1999), as well as earlier by Lakoff and Johnson (1980).

(4) The idea that "in natural language meaning consists in human interpretation of the world. It is subjective, it is anthropocentric, it reflects predominant cultural concerns and culture-specific modes of social interaction as much as any objective features of the world 'as such'" (Wierzbicka 1988: 2, reiterated by Wierzbicka 1992, and previously elaborated by Whorf: see Carroll, 1956).

In short, if forced to name the framework espoused in this monograph, I would (in a somewhat tongue-in-cheek fashion) call it "contrastive functional cognitive cross-cultural linguistics."

Within the present framework, lexical anisomorphism is defined as any divergence from full one-to-one systemic lexical equivalence – that is, the lack of any possible differences in any relevant linguistic features other than the physical form between the source-language lexeme and its target-language equivalent. The distinctive feature of this concept is thus rooted in contrastive linguistics, but its content draws from other aforementioned approaches. The manners in which divergence manifests itself can be based on different experiential and social realities, on different functioning and networking, and on different metaphorical extensions – all of which are central in systemic functional, cognitive, and cross-cultural linguistics.

It is important to have in mind that any case of lexical anisomorphism is always manifested in one direction (going from source language to target language), at a given moment in time (in a typical case, with a small exception of classical languages and historical bilingual dictionaries, we are not contrasting an ancient period of one language with a contemporary period of another), and at a particular proficiency level (it would not make sense to contrast the vocabulary of one language at the novice level of proficiency with that of another language at the superior level).

It should be emphasized quite strongly that the exploration of CLA here is restricted to the systemic level (or, in other words, the speaker-hearer's knowledge of the language: competence, in the sense of Noam Chomsky, 1965) – that is, to the vocabulary of the two contrasted languages. The textual level (that is, concrete utterances: performance, in the sense of Chomsky, 1965) remains outside the scope of analysis. This approach is thus differentiated from approaches such as the cognitive linguistic study of metaphor, in which the analysis is often performed at the textual level. The reason for concentrating on the systemic level is that all applied linguistic endeavors, lexicography in particular, strive to encapsulate this level of lexical functioning.

The present definition of CLA rests on two elusive concepts: that of language and that of the lexeme (or, in less technical terms, the word).

Any discussion of language at a cross-linguistic level must also include mention of dialect. The arbitrariness of the distinction between a language and a dialect is perhaps most colorfully illustrated by the well-known quip *A shprakh iz a dialekt mit an armey un flot* ("A language is a dialect with an army and navy"), most commonly attributed to Max Weinreich (1945), a scholar of Yiddish, considered a separate language by some and a dialect of German by others. It is indeed the case that no criteria exist to distinguish what a separate language is (as opposed to a dialectal or other variety of the same language). However, the decision whether to declare an entity one of the dialects or other varieties of a language rather than a separate language will hardly affect the nature of CLA. For example, if we compare a pair of equivalents between English (which can be British, American, Australian, etc.) and Serbo-Croatian (which can be Serbian, Croatian, Bosnian, etc.), the difference will still exist for all practical purposes, whether we consider each of the aforementioned varieties separate languages (as advocated, for example, by Mencken, 1919, for American English, Baker, 1945, for Australian English, and numerous authors for Bosnian, Croatian, and Serbian (BCS): see Kordić, 2004; 2008; 2010, for more information) or just varieties of one respective language. The same is true about other similar cases (Iberian and various forms of Latin American Spanish, Lusitanian and Brazilian Portuguese, German, Austrian, and Swiss German, etc.). Only the test of time can (in)validate the claims in this field, which can be seen when we read Sweet (1877: 196) today: "The result of these and similar changes will be that in another century any fixed scheme of reform adopted now will be nearly as unphonetic as our present Nomic spelling. It must be remembered that by that time England, America, and Australia will be speaking mutually unintelligible languages, owing to their independent changes of pronunciation."

Similarly difficult is the delineation of lexemes from non-lexemic entities. One can assume various strategies in this field but, for all practical purposes, we will consider lexemes those items one would normally look up in a dictionary: words, idioms, and lexical affixes (but not grammatical affixes, syntactic frames, collocations, etc.). These entities can engage in all forms of CLA to be discussed further in this text. In this regard, the distinction between a *lexical unit* and *lexeme*, introduced by Cruse (1986: 49, emphasis in original) may be helpful:

Lexical units are those form-meaning complexes with (relatively) stable and discrete semantic properties which stand in meaning relationships such as antonymy (e.g. *long : short*) and hyponymy (e.g. *dog : animal*) and which interact syntagmatically with the contexts in various ways... Lexemes, on the other hand, are the items listed in the lexicon, or "ideal dictionary" of a language.

The objects of research here, as was the case with Cruse (1986), are lexical units rather than lexemes.

It has to be admitted that the concepts of language and of lexeme are virtually indefinable, therefore forcing the adoption of practical solutions. Another problem is that this research is trying to extricate discrete categories from what is in fact an indiscrete continuum (and this applies to both the types of CLA to be established and the strategies for their treatment). I fully agree with Hartmann (2007b: 183), who states:

There is a cline of cultural diversity from the most general ("universal") to the most specific ("unique")...

There is a scale of equivalence from most complete ("full") through partial to most incomplete ("zero" or "nil")...

There is a range of translation methods from the most literal ("transfer" and "substitution") via contextual transposition ("modulation") to free adaptation ("circumlocution"). [. . .]

There is a set of lexicographical devices for presenting translation equivalents from text-insertible matches via labels and illustrations to explanatory glosses.

Having said all that, one should note that only discrete categories can be operational. Reducing an indiscrete scale to discrete points certainly has its limitations, but that is what we are forced to do in many other spheres of linguistics, for example in establishing and assigning the levels and sublevels of any given language proficiency scale.

On a related note, the practice-driven nature of the present monograph should be strongly emphasized. The need to encapsulate the patterns in which the lexical systems of any given two languages can differ stems from a set of practical challenges in applied linguistics. The cases of CLA are catalogized in Chapter 3, with an eye toward the applicability of the categories in applied linguistics endeavors. Among other things, this assumes that the categories are simple enough to be operational and deployable.

There are three major differences between the present research and relevant linguistic approaches (outlined in Chapter 2). First, while most approaches look for commonality in the languages of the world, the present research looks for the patterns in which they exhibit a difference. Second, in theoretical linguistics, cross-linguistic lexical anisomorphism is typically addressed either at a broader level, when the main claim is often about its unpredictability, or at the level of case studies of particular words. The current research is missing a construct that would fill the gap between the two aforementioned levels and offer a workable and deployable taxonomy of the phenomenon. Third, while some linguistic approaches claim that their models reflect psychological or neurological realities, the model of CLA presented here is meant to be a useful intellectual construct and nothing more than that. At no place in this monograph is it claimed that any of the categories of CLA are lodged in psychological or neurological realities. The

categories are arbitrary constructs, motivated and verified solely by the need to account for the broadest data set with the simplest possible tool.

The overarching question here, then, is if cross-linguistic lexical aniso-morphism lends itself to establishing simple yet comprehensive patterns of differences that will be useful in applied linguistics and related disciplines.

The phenomenon of full one-to-one equivalence, used as a building block for a definition by negation of the central concept in this monograph, is encoun-tered extremely rarely. As Zgusta (1971: 316) aptly observes:

We know how complex the lexical meaning is. Absolute equivalence requires, then, that the lexical meaning of the two lexical units be absolutely identical, in all components (designation, connotation, range of application). Because of the anisomorphism of languages..., such absolute equivalents are rather infrequent (outside the domain of scientific terminologies). The usual situation is that the lexical meaning of the respective lexical unit of the target languages is only partly identical with that of its counterpart in the source language.

One should say that, even in medical (and other scientific) terminology, only pairs such as (Eng.) *meningitis* versus (Serbo-Croatian) *meningitis* would be considered full equivalents, while pairs such as (Eng.) *appendicitis* versus (Serbo-Croatian) *upala slijepog crijeva* ("appendicitis," "appendix inflamma-tion"; lit. "inflammation of the blind intestine") would not, inasmuch as the latter term uses a mental image that is not present in the former. A number of similar Spanish–English anisomorphic pairs in terminology are given by Gómez González-Jover (2006), while Tarp (1995) offers several such examples between Danish on the one hand and French, Spanish, German, and English on the other. It is safe to say that the scope of the concept discussed in the present monograph covers the vast majority of lexical equivalents between any two languages.

Lexical anisomorphism is only a subset of general cross-linguistic aniso-morphism, but the practical needs of lexicography have given considerable prominence to this notion. Credit is definitely due to Ladislav Zgusta's ground-breaking *Manual of Lexicography*, in which he states that the principal task of a bilingual dictionary is the coordination of lexical units of the source language with that of the target language. He then goes on to say (Zgusta, 1971: 294):

The fundamental difficulty of such a co-ordination of lexical units is caused by the anisomorphism of languages, i.e. by the differences in the organization of designate in the individual languages and by other differences between languages.
    What leaps most to the attention of even the average layman are the cases of the so-called culture-bound words.

Furthermore, he points out (296):

It would, however, be completely wrong to limit the concept of anisomorphism and the discussion of it to the "culture-bound words" only. On the contrary, anisomorphism must be expected in all lexical units and can be found in most of them.

Indeed, even a cursory review of equivalents in any exhaustive bilingual dictionary would reveal instances of lexical anisomorphism such as zero equivalence (for example, the Chinese 叩头 [*kòutóu*] "knock head," the ritual bowing and knocking of the head on the ground in front of the emperor, does not have an equivalent in various other languages, and it needs to be described), multiple equivalence (found in the aforementioned situation that the English word *brother* has two Mandarin Chinese equivalents: 弟弟 [*dìdì*] and 哥哥 [*gēgē*]), different connotations (for example, when used about appliances, the Serbo-Croatian *crći*, "to die," is colloquial, while its English equivalent *to die*, as in *My car died*, remains neutral), different syntagmatic and paradigmatic links (German mustard is not "hot" but, rather, "sharp" – that is, *sharfer Senf* – while in Russian mustard cannot be "sharp," it has to be "strong" – that is, сильная горчица [sil'naja gorčica] – although peppers *are* "sharp": острый перец [ostryj perec]), different frequency (both Arabic words for *mosque*, viz. مسجد [*mašžid*] and جامع [*ža:mi'*] are far more frequent in that language than their English counterpart), different lexical structuring (Eng. *peephole* is a compound term that invokes the frame of peeping through a hole, as does the German *Guckloch*, which means roughly the same, while its Polish equivalent, *judasz*, literally "Judas," and the less common English synonym *Judas(-hole)* bring about an elaborate Bible-based metaphor, but the German synonym *Spion*, "spy," invokes a less elaborate metaphor), and many others. Each of these situations has its applied linguistic consequences. To take an obvious example, in compiling a bilingual English–Mandarin Chinese production dictionary, one would need to include a gloss that the first Mandarin equivalent of the English word *brother* refers to one's younger brother and the second one to one's older brother; something along the lines of "brother [. . .] 1.弟弟 *[dìdì]* (younger brother); 2. 哥哥 *[gēgē]* (older brother)."

My discussion of lexical anisomorphism will adopt the following agenda. Its theoretical goal is to discuss the scope and the internal structure of the phenomenon and its further theoretical consequences. Its applied linguistic goal is to propose solutions to the problems caused by lexical anisomorphism in lexicography. Its two overarching principles are to describe the phenomenon as succinctly as possible and to find the simplest solutions possible. The two goals, theoretical and applied linguistic, are interrelated. On the one hand, theoretical distinctions should serve as the basis for applied solutions, and, on the other hand, applied linguistic problems point to theoretical distinctions. In this regard, it should be emphasized that the primary goal is to establish deployable, operational categories. The present research is strongly practice-driven.

It is often the case that theoretical linguistic categories tend to be overly intricate (sometimes even to the point that the explanation is more complicated than the phenomenon that is supposed to be explained). This then makes them non-operational in any applied linguistic endeavors. The present research is

conducted with an eye toward bridging the gap between the two principal branches of linguistics (theoretical and applied) by designing a theoretical taxonomy simple enough to be operational in applied linguistics. In other words, the goal is to come up with a minimal set of categories that will be exhaustive enough to cover the field of CLA.

In this spirit, Part I, the first step of the aforementioned agenda, represents an attempt to build a comprehensive yet succinct and operational taxonomy of cross-linguistic lexical differences. It is important to stress that the envisaged taxonomy is meant to be a heuristic rather than a deterministic mechanism. Heuristics are more flexible in accommodating complex and unpredictable data systems such as CLA, and at the same time they are much more deployable then their deterministic counterparts. For all practical purposes of various applied linguistic tasks, a simpler explanation that works in the vast majority of cases is better than an explanation, which may be absolutely accurate but too complex to be deployable.