ORIGINAL PAPER

When MIPVU goes to no man's land: a new language resource for hybrid, morpheme-based metaphor identification in Hungarian

Gábor Simon[1] · Tímea Bajzát[2] · Júlia Ballagó[3,4] · Zsuzsanna Havasi[2] ·
Emese K. Molnár[2] · Eszter Szlávich[5]

Abstract

The aim of the article is to present a new language resource for metaphor analysis in corpora that is (i) a MIPVU-inspired, morpheme-based process for identifying metaphor in Hungarian and (ii) the refinement and innovative version of metaphor identification extending the scope of the process to multi-word expressions. The elaboration of language-specific protocols in metaphor identification has become one of the central endeavors in contemporary cross-linguistic research on metaphor, but there is a gap in the field regarding languages with rich morphology, especially in the case of Hungarian. To fill this gap, we developed a hybrid, morpheme-based version of the original method, which can handle morphologically complex metaphorical expressions. Additional innovations of our protocol are the measurement and tagging of idiomaticity in metaphors based on collocation analysis and the identification of semantic relationships between the components of metaphorical expressions. The present paper discusses both the theoretical motivation and the practical details of the adapted method for metaphor identification. As a conclusion, the presented protocol can provide new answers to the questions of metaphor identification in languages with rich morphology and shed new light on the internal semantic organization of linguistic metaphors.

Keywords Metaphor identification · Morpheme-based MIPVU · Semantic relations · Idiomaticity

✉ Gábor Simon
  simon.gabor@btk.elte.hu

1 Department of Modern Hungarian Language, Eötvös Loránd University, Budapest, Hungary

2 Eötvös Loránd University, Budapest, Hungary

3 University of Debrecen, Debrecen, Hungary

4 Hun-Ren Hungarian Research Centre for Linguistics, Budapest, Hungary

5 Department of Digital Humanities, Eötvös Loránd University, Budapest, Hungary

∅ Springer

1 Introduction

According to the Macmillan Dictionary,[1] the expression *no man's land* has a military meaning as its basic sense ('an area of land that separates two armies'). In the title of this study, the contextual meaning of the expression is obviously not the same; rather, it is a more general and abstract sense provided as a second meaning in the dictionary: 'a situation that is not easy to deal with because it is not clear what rules or policies are operating' (analyzing the polyword as a compound noun, i.e., a single lexical unit, see Steen et al., 2010, pp. 30−31). Therefore, according to the process of metaphor identification by the MIPVU (Metaphor Identificátion Procedure Vrije Universiteit) protocol (Steen et al., 2010), the expression in the title of the present paper can be considered metaphorical and can be marked as MRW, or Metaphor Related Word, since it has the potentiality of metaphorical interpretation.

The basic meaning of the corresponding Hungarian expression *senki földje* (nobody land-poss3sg, 'no man's land')[2] is given in the same way by The Concise Dictionary of Hungarian (second edition, Pusztai, 2003, henceforth CDH). The contextual meaning in the Hungarian version of the title is, however, slightly different from the previous one: the dictionary lists 'unowned land' as the second sense. Dealing with metaphor identification in Hungarian confronts the annotator with both interpretations. Since systematic metaphor identification in Hungarian has no research precedents, this language can be considered *no man's land* for the MIPVU method. Moreover, because of the rich morphology of the language, adapting MIPVU to Hungarian is not without significant challenges for the researcher. Even though the method is well-developed, language-specific issues (for example verbal prefixation or the metaphorical meaning of case markers) raise difficult questions in which it is not clear whether the annotator should closely follow the principles of the original protocol. And although several improvements have been made as a result of adjusting MIPVU to agglutinative languages (e.g., Polish, Lithuanian, Serbian, and so on), a comprehensive approach to morphologically complex structures and idiomatic expressions is yet to be invented.

The aim of the present paper is twofold: firstly, to adapt MIPVU to Hungarian (a language with rich morphology), and secondly, to widen its scope by identifying and annotating semantic relationships between the components of multi-word metaphorical expressions. Crucially, the proposed adaptation is not confined to minor amendments but involves introducing profound changes in the methodology. In other words, this study presents an alternative, morpheme-based hybrid process of metaphor identification (which is called MetaID protocol after the name of our research group),[3] relying fundamentally on the principles of MIPVU but fine-tuning

---

[1] https://www.macmillandictionary.com/dictionary/british/no-man-s-land (last access: 25th May 2022).

[2] In the study, the Leipzig Glossing Rules are adopted to mark the exact linguistic structure of the Hungarian expressions. For more information, the reader can visit the following URL-page: https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf (last access: 11/30/2022).

[3] The MetaID Research Group is a team of graduate and post-graduate students working on the issue of metaphor identification in Hungarian. The group was set up and is led by the corresponding author of the present paper, at the DiAGram Research Centre in Functional Linguistics at Eötvös Loránd University Budapest.

and extending them with the purpose of systematically handling morphological structures, idiomaticity and semantic relations within complex metaphorical expressions. The main aim of the proposed innovations is to enhance the precision of metaphor analysis not only in a specific language but also at a more general level.

The need for applying MIPVU to languages other than English does not come, of course, out of the blue. This was one of the promising perspectives emerging after the publication of the protocol: as a reviewer put it in 2014, "further research should systematically seek to apply the method to other languages" (Pérez-Sobrino, 2014, p. 144). Fortunately, this is no longer a mere promise: those interested in metaphor identification in a wide range of languages (from French to Serbian, from Scandinavian languages to Uzbek, and from Chinese to Sesotho, to mention only a few) can turn to a collection of case studies published by Nacey et al. (2019, see also Lu and Wang, 2017 for Mandarin Chinese adaptation). The scope of the volume extends to numerous language families (including Indo-European, Sino-Tibetan, Niger-Congo and Turkic) and different types of languages (e.g., with or without rich morphology).

Hungarian could be of particular interest for MIPVU and metaphor identification in general because of its peculiar grammatical system and flexible patterns of word formation, which—according to the editors of the volume cited above—"present challenges" for the method (see Nacey et al., 2019 with special reference to Lithuanian and Polish). However, Nacey et al. (2019) make no mention of any research on Hungarian or indeed any other Finno-Ugric language. (This is hardly surprising as the book is intended as a "solid start" rather than a representative handbook.) And even when we look beyond the MIPVU protocol, corpus-based investigations of metaphor in Hungarian turn out to be very limited in scope. Although one can find more or less corpus-assisted studies on metaphorical expressions in Hungarian (e.g., the cross-linguistic analysis of blood in American English and Hungarian by Simó (2011), studying the metaphorization of body parts in English, German and Hungarian (Tóth-Czifra, 2014), or assessing the processes of automatic metaphor identification in Hungarian (Babarczy et al., 2010), they are less systematic from a methodological point of view. Overall, it seems fair to say that Hungarian has remained a relatively understudied language in the field of corpus-based analysis of metaphorical expressions.

We want to change the terra incognita status of Hungarian (cf. the geographical analogy in the title of the study), putting it on the map of metaphor identification, as well as expanding the scope of the MIPVU analysis to phenomena being strongly related to metaphorization but lacking a systematic treatment in the framework of the original method. The benefits of such an effort are twofold. It confronts the original protocol with new phenomena, also motivating a useful extension of it. Moreover, it results in a hybrid method of metaphor identification integrating MIPVU with a rigorous analysis of morphological structures and construction-internal semantic relations which aims at shedding new light on patterns of metaphoric use in a particular language.

Our paper aims to provide a solid theoretical foundation for the MetaID protocol, and then it gives a detailed overview of it. The primary focus is on theoretical and methodological questions rather than empirical results. Although our method proved to be reliable in a testing phase, supporting the recent large-scale annotation of a

research corpus, adapting the original protocol by refining it towards a hybrid process of identification presents a range of issues for further discussion. For this reason, the present paper only addresses the theoretical and methodological problems of adaptation, leaving the presentation of empirical results for a later study.

The paper is structured as follows. After discussing the theoretical problems of identifying metaphorical meaning in Hungarian expressions, we present our innovations in three subsections: modifications essential for adapting MIPVU to an agglutinative language (Sect. 2), additional changes that extend the analytical scope of the identification process (Sect. 3.1), and finally, the systematic presentation of the extended tag set of the annotation (Sect. 3.2, with subsections for each tag). Following the discussion of relational labels of metaphoricity used in the process of identification (Sect. 3.3), the guidelines of the MetaID protocol will be detailed, too (Sect. 3.4). The paper briefly discusses the design of annotation regarding the dictionaries and the tools for digital language analysis used in the elaboration of our hybrid protocol (Sect. 4.1) and provides preliminary testing of the reliability of annotation (Sect. 4.2). Finally, it ends with concluding remarks and an outline of future perspectives (Sect. 5).

2 Theoretical background: issues of adapting MIPVU to Hungarian

Theoretical and methodological innovations go hand in hand during the process of adapting a protocol to a new language: Nacey and her colleagues also openly admit that "when it comes to the more nuanced details of the protocol, researchers found themselves struggling to adjust the method in ways that would address language-specific issues" (Nacey et al., 2019, p. 2). They categorize the problems emerging through adaptation as operational issues and language-specific features. Operational issues include choosing an adequate dictionary (see Kaya, 2019, p. 229; Veale et al., 2016, p. 63), carefully determining the basic unit of analysis or tackling the fuzzy boundaries of metaphor. By contrast, analyzing such language-specific features as the meaning of compounds, the semantics of inflectional morphology, case markers and postpositions, and handling the flexible grammatical behavior of preverbs require not only refinement but also partial modification of the original method.

To give a brief overview: with an eye to ensuring a high level of accuracy in the analysis of metaphorical expressions in Hungarian, essential modifications of the protocol were necessary during the adaptation process. Some of them belong to the language-specific issues mentioned above: (i) carrying out morpheme-based rather than word-based annotation; (ii) elaborating a careful analysis of the components of compounds; (iii) dealing with preverbs and inflectional morphemes as metaphor-related components. In this section, all these modifications are discussed in detail.

Adapting an existing annotation schema to a language different from the original (i.e., to which the schema was first implemented), does not necessarily mean applying the same annotation tags for labelling the "same" phenomena. As each language is different in how it conventionally symbolizes intersubjective human experience, a careful adaptation of MIPVU must be responsive to the specificities of Hungarian. Therefore, we had to adjust the annotation protocol to the basic typological features

of Finno–Ugric languages with no existing example concerning members of this language family.

The core challenge we had to face was the agglutinative character of Hungarian, which allows metaphorization to emerge at a sub-token level. For example, the nominal inessive case marker *-ban/-ben* with the basic meaning 'inside of ‹something›' may prompt for cross-domain mapping in certain contexts: *a könyv-ben* the book-ine 'in the book', *fej-ben* (*tart*) head-ine (keep) '(to keep it) in the head', or *eb-ben a hónap-ban* this-ine the month-ine 'in this month'. In these expressions, the case marker invites construing the noun as referring to an (abstract) container. As can be seen, the Hungarian case marker functions similarly to the English preposition *in*, and since prepositions are handled consistently in MIPVU, we could adopt the steps of analyzing them as proposed by Steen et al. (2010, pp. 28–30, 98–101). However, in Hungarian (as in other agglutinative languages) the morphological case marker is not separable from the noun, therefore we cannot treat it as a metaphorical lexical unit even though its contextual meaning would definitely support this categorization. Put simply, the metaphorization of grammatical cases is "invisible on the lexical level of analysis" (Bogetić et al., 2019, p. 214).

This problem, of course, is not peculiar to Hungarian, although researchers of other languages had to face up to sub-token metaphorization in different ways. In Polish, the system of prefixes and the vocative case constitute the morphological level of metaphorization, but the authors propose only some specific rules for treating these phenomena, and they mark the whole formation as metaphorical (Marhula & Rosiński, 2019). In Lithuanian, prefixes and nominal cases raise difficulties for the annotators, and the proposed solution is similar: in a further step, the annotator needs to investigate the morphological structure of the word form, identifying the basic meaning of the prefix (or case marker) and the base word, then according to the contextual meaning the whole formation can be annotated as metaphorical (Urbonaité et al., 2019). In the case of Sesotho (Seepheephe et al., 2019), prefixes and the ambiguous status of orthographic words (they do not receive PoS tagging necessarily) cause the majority of problems in demarcating lexical units and result in a long list of exceptions of a word-by-word analysis of lexical units. And although Kaya (2019) recommends that researchers of the Uzbek language follow the original MIP (Metaphor Identification Procedure) method and detect metaphoricity at the level of the entire word, the question of whether metaphorical meaning "derives solely from the case suffix, or from the word form as a whole" (Kaya, 2019, p. 237) remains partly open.

The adaptation of MIPVU to Serbian has gone the farthest in this direction so far: Bogetić and her colleagues encountered the problems of compounding and grammatical case marking, on the grounds of which they propose a "case-sensitive MIPVU" (Bogetić et al., 2019). Their protocol includes an additional step: the analysis of the grammatical case and its contextual meaning. This step allows us "to identify case-encoded metaphoricity, while keeping the analysis grounded in the lexical level" (Bogetić et al., 2019, p. 215), which means that the Serbian protocol makes it possible to distinguish lexical metaphor from inflectional metaphor. As a conclusion, rich morphology appears to raise serious problems for metaphor identification, but the solution to these problems in most of the existing adaptations of MIPVU is

limited to exploring the morphological structure of the lexical unit and marking the potential of metaphorical meaning of the word as a whole. The most radical innovation appears to be the annotation of inflectional metaphors in Serbian; nevertheless, a comprehensive account of morphological structures in metaphor identification is yet to be elaborated.

As a further, but well-established step toward an extended method of annotating metaphorical expressions, the MetaID protocol treats morphemes rather than lexical elements as the basic units of analysis. In other words, instead of building an annotation schema on lexical units, our method applies a morpheme-based annotation protocol.[4] This implies that each morpheme identified by the analyzer (the e-magyar NLP tool in our case) may receive an annotation tag and a novel label has been introduced for metaphorical inflections: MKI, i.e., 'metaphor-related inflection'.

Considering the new unit of analysis, we laid some principles concerning morphemes potentially receiving metaphor-related annotation tags. On the one hand, we had to introduce some stipulations in accordance with the basic purposes of metaphor identification, i.e., adjusting annotated phenomena to everyday language users' knowledge. Thus, regarding word forms produced by derivational affixes, the only affixes that may receive tags are the ones that do not change the word class of a given word form. This is motivated by the principles that (i) the method does not strive to detect etymological aspects of metaphorization, and (ii) conversion can influence meaning extension, therefore a comparison between the basic and the contextual meaning of the same lemma instantiating different word classes is not possible (see Steen et al., 2010, pp. 33–36).

On the other hand, a morpheme-based annotation protocol needs to deal with the issue of compound words, which requires careful delineation of the annotation principles. Concerning compound words, the MetaID annotation schema follows the principle of lexicalization which is determined on the basis of dictionaries (see also the case of the Sesotho language, Seepheephe et al., 2019). If a compound word (e.g., *vér-nyomás* 'blood pressure', *jel-szó* lit. sign-word 'slogan/password') can be found in CDH as an entry word, particular components of the compound do not receive metaphor-related tags in their own right. However, in cases when a compound word does not seem to be lexicalized according to the dictionary, the protocol reserves the possibility for particular components to receive a metaphor-related annotation tag (e.g., *válasz-forma* lit. answer-form 'form of answer').

Switching over to a morpheme-based protocol not only affects derivational affixes and compound words but also allows for the annotation of metaphorization associated with preverbs and inflectional morphemes (e.g., case markers). In Hungarian, preverbs often have a basic directional meaning that may support various cross-domain mappings. For example, in the construction *le-csendesít-i a*

---

[4] Although this decision seems to be a profound modification of the original protocol, it cannot be claimed that there are no precedents of focusing on morphological structures in MIPVU or in the broader endeavor of metaphor identification in general. Steen (2007, p. 50) and Steen et al. (2010, p. 41) explicitly argue that morphemes can initiate metaphorical meaning. Nevertheless, in MIPVU morphological units are rejected as the units of analysis (see Steen et al., 2010, pp. 12−13, 26−27).

vita-partner-ek-et (down.prev-calm-prs.3sg the discussion-partner-pl-acc 'calms down discussion partners'), the preverb *le* means 'downwards' according to CDH, but in the given construction, this directionality in the meaning of *le* 'down' can be said to undergo metaphorization (producing the contextual meaning of 'causing a decrease' and 'getting into a calm state of mind'). Thus, in cases illustrated by the example, preverbs (which may also occur in reversed and interrupted word order with respect to the associated verbs) can be tagged as "metaphor-related expressions" (abbreviated as MKK in our tag set) on their own right as morphemes participating in cross-domain mappings. However, metaphorization in the above construction does not emerge independently from the other components of the construction; rather, the word stem also has a role—albeit only secondarily—in the emergence of metaphorical meaning (the dictionary-based contextual meaning of the preverb refers to calmness as well, originating from other conventional expressions such as *lehiggad* 'calm down' and *lenyugszik* 'get relaxed'). Accordingly, we introduced a new tag ("new" is compared to the original MIPVU method) called "metaphor-related component" (abbreviated as MKKomp) reserved for those morphemes that take part in the emergence of metaphorical meaning of a construction, although they are not the initiators of it, in other words, they do not play a "triggering role" in the process.

Regarding the initiation of metaphorization, Hungarian inflectional morphemes may show a behavior very similar to preverbs. Embedded in certain bigger constructions, they may have the potential to initiate cross-domain mappings. For instance, in the construction *ez-en a szomorú nap-on* (this-sup the sad day-sup 'on this sad day'), the nominal superessive case marker *-on* 'on' means 'being on the top of something' according to the baseline, but in this construction—linked to the noun *nap* 'day'—it conceptualizes a given unit of time as a surface, thus, it initiates cross-domain mapping (similarly to the case of *over* in English, see Steen et al., 2010, p. 111). For labelling this type of metaphorical meaning generation, we added the "metaphor-related inflection" (MKI) tag to the annotation tag set. When this tag is assigned to an inflectional morpheme, the word stem is tagged as a "metaphor-related component" (MKKomp), since—analogously to the verb in the preverb + verb composite structure—the nominal stem is also influenced by the metaphorization that is prompted by the inflection.

In our approach, the annotator can dispense with morphological analysis only in one case. If the inflected nominal form belongs to a verb as its argument, and the inflection is used as a conventional case marker being obligatory in that specific construction, then the whole nominal form is allocated a specific tag ("metaphor-related argument" or MKA in our set), without marking any metaphorization on the inflection itself (and labelling the nominal stem as a component of figurative meaning generation). As an example, the construction *beszél valami-ről* (talk.3sg something-del 'talks about something') has a nominal argument in the delative case referring to the topic of the talking process, but in this case, the inflection does not have any directional meaning. In other words, its basic spatial meaning ('away and down from') is not active in the construction from the perspective of the speaker, and hence it cannot function as a trigger of metaphorical interpretation. Note that Steen

et al. (2010, pp. 29−30) offer a similar solution to handle the differences between prepositional and phrasal verbs. (For further details see our guidelines in Sect. 3.4)

This section summarized the basic theoretical reorientation of the MIPVU protocol towards morphologically complex linguistic structures, as well as the most fundamental consequences resulting from this decision. It seems plausible to claim that these changes allow for an extension of the original method to agglutinative languages established by other authors, making a broader cross-linguistic application of the identification method possible. There are, however, further innovations following the aim of improving the method's accuracy. The next section brings these into focus.

3 From MIPVU to the MetaID method: additional innovations and a hybrid methodology for metaphor identification

There are further extensions of the method in our research, which are not motivated by the process of adaptation but rather by the more general purpose of making metaphor identification as precise as possible. As Steen declared in an interview (Nacey et al., 2019, p. 10), the aim of the MIPVU protocol is "to measure the linguistic manifestation of metaphor as a cross-domain mapping on thought, largely in the tradition of cognitive linguistics". Taking this aim of measurement seriously, our research group has completed the original process by marking the semantic relations between the components of a metaphorical expression (using the terms of Cognitive Grammar, see Langacker, 2008, 2013) and indicating the idiomaticity of a metaphor related expression (relying on the measurement of collocational behavior in a reference corpus).

With these innovations, metaphorization patterns can be explored on a more precise level than has been achieved by any previous analysis: not only is it possible to identify potentially metaphorical expressions with our protocol but also their schematic organization as composite structures can be observed. Moreover, levels of conventionality can be assessed as well. Hence, our analysis goes beyond the recognition of metaphorical lexical units by also supporting the qualitative investigation and quantitative measurement of the internal organization of complex units of metaphorization. In this area, however, our hybrid methodology is still fundamentally inspired by the MIPVU protocol: it adopts most of the original tags and builds the procedure on the steps proposed by Steen and his colleagues. Thus, innovative alterations of the model are actually aimed at increasing the method's efficacy in investigating complex metaphorical patterns.

The aspects of our hybrid identification procedure dealt with in the present section are the following: (i) avoiding the use of the WIDLII ('When In Doubt, Leave It In') category; (ii) developing a method for identifying and analyzing idiomatic expressions on the basis of their collocational behavior; (iii) creating a new tag set for identifying and annotating semantic relations between components of a metaphorical expression. (The last group of innovations concerning annotation design such as (iv) choosing an appropriate dictionary for the process; (v) using further

devices for carrying out tokenization and collocation measurement; (vi) implementing the annotation process with online collaboration are detailed in Sect. 4)

3.1 Theoretical refinements and qualitative exploration of metaphorical expressions

Metaphor identification—in every language and on every level of precision—leads the annotator to hard decisions due to the fact that there is a fuzzy boundary between metaphor and other kinds of indirectness in meaning generation (e.g., metonymy or generalization). While MIPVU follows a more inclusive way of dealing with these problematic cases, using the WIDLII category for non-ambiguous expressions, in developing the MetaID protocol we agreed upon refining the available protocol and proposing a rather exclusive way of tagging metaphorical expressions. Therefore, while we relied again on the basic principles of MIPVU (e.g., not considering metonymy or generalization as potentially metaphorical in their nature), additional efforts have been made to sharpen the procedure's focus on metaphor. The main innovation of our procedure in this respect is not using the WIDLII category.

This category ensures the flexibility of operationalizing the notion of metaphor (including borderline cases of metonymical meaning extensions and avoiding a rigid distinction between metaphorical and non-metaphorical data), nevertheless, it has the risk of producing an eclectic type of data (Pérez-Sobrino, 2014, pp. 142−143). Moreover, using this category may result in a more heterogeneous set of data annotated as metaphors, increasing the internal variation of metaphor without reflecting any qualitative differences. In other words, using WIDLII may increase the recall of the analysis, but it also decreases its precision. In our workflow, careful curation is an essential step in implementing the protocol, and since it results in the unanimous labelling of the problematic cases (after a discussion), WIDLII did not prove to be necessary. (This does not mean, of course, that the category itself has lost its motivation, or that WIDLII would not deserve more scholarly attention in the future).

To give an illustrative example, take a look at the following expression: *meredek lépcső visz le a homály-ba* (steep staircase take.prs.3sg down shadow-ill 'a steep staircase leads down into the shadow'). Although the example could be interpreted as personification on the grounds of the basic meaning of the verb *le-visz* (down-take 'lead down'), with the staircase being conceptualized as a human agent performing an act of taking somebody down (see Dorst et al., 2011 for the categories of personification), it is not the staircase which takes us down but our legs. Therefore, the meaning of the expression can be motivated solely by The place stands for the process metonymy. Since metonymical and personifying readings do not necessarily exclude each other in the process of conceptualization (see Dorst et al., 2011 for further discussion), the Hungarian expression could be categorized as a WIDLII-case. However, since the annotator does not have any device to decide whether alternative explanations are plausible enough, these ambiguous cases may increase the amount of false positive labels. To make the procedure more precise, we need as many true positive cases as possible. In other words, using or avoiding WIDLII is in tight relationship with the overall purpose of the research: obtaining as much data as possible

(and running the risk of a messy set of data) or obtaining exactly the expressions that fulfill the criteria of querying (without any additional sorting of the hits, see Sass, 2022). We firmly believe that in the current phase of cognitive and corpus linguistic research on metaphor the researcher needs a high precision on the level of semantics and a high recall on the level of linguistic pattern. While the former supports generalizations on the linguistic symbolization of true metaphorical meaning (i.e., qualitative explorations), the latter makes it possible to observe the linguistic variation of metaphor (i.e., quantitative analysis) as extended as possible (see Meurers & Müller, 2009, p. 922). Therefore, when the contextual meaning of an expression can be fully explained by metonymic conceptualization, the token in question is not tagged as metaphorical in MetaID protocol.

To decide whether 'problematic cases' can be tagged as metaphors or not, the annotator can rely on the dictionary definition of the basic and the contextual meaning. Consider the expression *A harag* […] *egészséges-nek is nevez-het-ő ember-i érzelem* (The anger […] healthy-dat also call-pot-presp human-adj emotion 'Anger […] may also be considered a healthy human emotion').[5] The adjective *egészséges* ('healthy') seems to be metaphorical (motivated by the psychological is physiological cross-domain mapping). Its basic meaning in CDH is 'being in good health, intact, not ill, or characteristic of such a person or organism', while the contextual meaning given in the dictionary is 'meeting the requirements (in its effects), appropriate'. Although anger is a psychological state, and the basic meaning of the adjective depicts a physiological state, the contextual usage of the latter describes anger not as a body-related phenomenon (which would be unambiguously metaphorical) but as appropriate in a given situation. To conclude: a careful consideration of the meaning descriptions (provided by the dictionary) and the usage or reference of the terms in context makes it possible to come to a yes/no decision about the metaphorical nature of the expression's actual meaning. In this case, anger as a healthy ('appropriate') emotion is rather a generalization (from a narrower conceptual field of the body to the more general field of human behavior). Needless to say, not using WIDLII requires a thorough process of curation of the individual annotations.

In addition to drawing a more prominent line between metaphor and other phenomena of indirect meaning generation (e.g., metonymy and generalization) and hence narrowing down the operationalized category of metaphor to exact patterns of cross-domain mappings, ensuring the accuracy of analysis also involves further challenges. One of them is the exploration of semantic relations within complex metaphorical expressions, the other is tackling idiomaticity. Our hybrid method offers a solution for both of these issues.

Concerning the construction-internal relations of metaphorization, the aim of the MetaID protocol is to highlight four basic and two technical semantic operations between the components of metaphors, relying on the framework of Cognitive Grammar (Langacker, 2008, 2013). The first two of these links are the elaboration of the primary figure of a schematic process (i.e., the trajector, which also gives

---

[5] In the glossing, pot refers to the derivative suffix meaning potentiality in Hungarian, while presp stands for the derivative suffix of the present participle form.

the name of the relation and its abbreviation: tr), and the elaboration of its secondary figure (i.e., its landmark, henceforth lm). These relation types can be applied not only to verb + argument structures but also to attributive constructions with an attribute derived from a verb, and to the internal semantic link between the components of a metaphorical compound. Further semantic relations in our approach include the possessive relation (poss) and an underspecified relation of semantic elaboration (ela) which connects an inflection to its stem or a preverbs to its verb. There are two additional technical links as well: the expm tag is used to designate the explicative relation between a metaphor flag (the signal of a directly expressed metaphor) and a direct metaphor, while the "r" tag is applied whenever two separable, but tightly related components cannot be tagged together (because of inversion or word order conventions). With the application of these novel relational tags, the method of metaphor identification extends to exploring the internal semantic structure of the expressions that may lead the annotator to extract construction-like patterns of metaphorization. (For further details and examples see the extended tag set of our protocol in Sect. 3.2)

With regard to the issue of idioms, previous research found a large overlap between metaphors and idiomatic expressions (see Moon, 1998 for a comprehensive discussion). On the one hand, metaphorization is one of the crucial factors behind non-compositionality in idioms, making the vast majority of fixed expressions metaphoric in their meaning. On the other hand, linguistic metaphors form a continuum between single-word expressions to multi-word units of language (see also Deignan, 2005, p. 155 for the idiom-like linguistic fixedness of particular metaphors). Therefore, measuring the idiomatic nature of metaphor-related words provides a further step toward extracting metaphorical constructions from a corpus. Moreover, it can be considered an additional factor in estimating the conventionality of metaphorization.

While the careful distinction of metonymy or generalization from metaphor depends partly on the interpretation of the annotator, we made a significant attempt to make the analysis of idiomaticity as transparent as possible in detecting idiomaticity during metaphor identification. We also admit, of course, that intuition cannot be (and, indeed, need not be) ignored in the process of annotation and metaphor analysis in general. Intuition always plays an important role in detecting potentially metaphorically used expressions (even the term "potentially" refers to the fact that the result of the analysis is not a general truth but a possibility of metaphorical interpretation instantiated by the individual language user relying on their intuition). However, to maintain the replicability of the annotation, it is transparency that needs to be ensured. In order to properly identify the idiomatic character of metaphor-related expressions, the MetaID protocol introduced specific measurements. More specifically, when a given construction is already identified as metaphorical relying on the dictionary-based semantic disambiguation method, and it seems to be "idiomatic" (i.e., members of the construction co-occur extremely frequently) according to the intuition of a professional annotator, the protocol provides the annotator with additional support to validate this intuition: it requires a measurement of the collocational behavior of constructional components. For assessing the strength of collocation, the method relies on a corpus query tool working on the Hungarian Web 2012 (huTenTen12, for the Tenten corpus family, see Jakubícek et al., 2013) corpus.

When the measured collocational index is above a predetermined threshold (for methodological details, see 4), the main component of the construction triggering metaphorization receives the "metaphor-related idiomatic expression" tag (MKKid in Hungarian), whereas the other, non-central component of the metaphoric structure receives "metaphor related idiomatic argument" tag (abbreviated as MKAid) (in a verb + argument structure), or "metaphor related idiomatic component" tag (MKKompid in short) in a non-verbal idiomatic structure.

To sum up the most important additional innovations of our protocol, we can claim that it makes a more precise identification of metaphorical expressions possible. Although we acknowledge the fuzzy boundaries between conceptual metaphor, conceptual metonymy and generalization, the MetaID method strives to annotate only true metaphorical potentiality on the linguistic level, without maintaining the WIDLII category for the dubious cases. Moreover, it explores the internal semantic organization as well as the prefabricatedness of multi-word units in a corpus, extending the scope of the original method to the constructional aspect of metaphorization. It is worth stressing that on the one hand—according to our intention— the innovations detailed here may contribute to the accuracy of metaphor identification, increasing its precision in terms of the semantics of the identified phenomena and extending its scope regarding the linguistic patterns of metaphorization. On the other hand (and strictly following from the previous argument), the hybridity of the MetaID protocol cannot be considered necessary to solve the problem of metaphor annotation in Hungarian (or in other languages with rich morphology); rather, it provides the researcher with a more fine-grained analysis of metaphorization in a given language. By way of explanation, if the linguist is interested in metaphorical patterns on a coarser level of analysis, the core features of the Hungarian MIPVU (namely, the morpheme-based analysis or the way it handles preverbs, postpositions or compounds) are enough to carry out a basic-level annotation. However, contemporary corpus-assisted metaphor research needs further innovation to gain a better picture of linguistic metaphor, and to provide an accurately annotated corpus for researchers. Consequently, our hybrid method integrates a language-specific adaptation and a comprehensive refinement of metaphor identification procedures available in the field. Having introduced the ways in which the MIPVU protocol has been modified and improved, with the aim of giving a holistic picture of our hybrid identification method, we now turn to the details of annotation, demonstrating the use of adopted and new tags in the next section.

3.2 The extended tag set for identifying metaphorical structures in Hungarian

The core group of labels designating metaphorical potential (the mtags group) is constituted by two types of markers. In the first subgroup (Sects.3.2.1−3.2.4), there are labels adopted from the original protocol with the same (or almost identical) function (MKK, dMKK, MZ, MKKimp constitute this group). In the second subgroup (Sects. 3.2.5−3.2.10), there are novel tags annotating inflections, arguments of a verbal construction and the idiomatic equivalents of the adopted and recently

Table 1 The tag set of metaphor-related expressions in the MetaID and the MIPVU protocol

Function of the MetaID-tag	Corresponding tag in MIPVU
Metaphor-related expression (MKK, Metaforához kapcsolódó kifejezés)	Metaphor-related word (MRW)
Metaphor-related inflection (MKI, Metaforához kapcsolódó inflexió)	Metaphor-related word (MRW) (or WIDLII)
Metaphor-related argument (MKA, Metaforához kapcsolódó argumentum)	−
Metaphor-related component (MKKomp, Metaforához kapcsolódó komponens)	−
Metaphor-related idiomatic expression (MKKid, Metaforához kapcsolódó idiomatikus kifejezés)	−
Metaphor-related idiomatic argument (MKAid, Metaforához kapcsolódó idiomatikus argumentum)	−
Metaphor-related idiomatic component (MKKompid, Metaforához kapcsolódó idiomatikus komponens)	−
Direct metaphor-related expression (dMKK, direkt Metaforához kapcsolódó kifejezés)	Direct metaphor (MRW, direct)
Metaphor flag (MZ, Metaforazászló)	Metaphor flag (MFlag)
Implicit metaphor-related expression (MKKimp, Metaforához kapcsolódó implicit kifejezés)	Implicit metaphor (MRW, implicit)

introduced labels (such as MKA, MKI, MKKid or MKAid). In the following, both groups are illustrated with examples (for an overview of the tags see Table 1).

3.2.1 MKK (metaforához kapcsolódó kifejezés, 'metaphor-related expression')

The MKK label denotes an expression for which there is a difference between the basic meaning of the expression given by the dictionary and the actual, contextual meaning, and there is some similarity or cross-domain mapping between the two. For instance, in the following construction: *a bennünk-et ér-t csapás* (the us-acc happen-pst.ptcp blow 'the blow that happened to us') the word *csapás* ('blow') is given the label MKK because its primary meaning as defined in the dictionary is 'strong blow', but in the context, it has the contextual meaning of 'serious trouble, affliction', which is the third meaning of the Hungarian word for 'blow' according to the dictionary.

3.2.2 dMKK (direkt metaforához kapcsolódó kifejezés, 'direct metaphor-related expression')

When a component of an expression (or another orthographic symbol such as "") is directly used to set up a cross-domain mapping, the token is marked as a direct metaphor-related expression. An example of this is the construction *akvárium-szerű-en ki-világít-ott üveg-kalicka* (aquarium-like-sup out.prev-light-pst.ptcp glass-cage 'a glass cage being illuminated as an aquarium') where the nominal stem *akvárium* prompt for a direct comparison of the topic of the discourse to an aquarium (while

the component -*szerű* 'like' makes this mapping explicit). Thus, it can be tagged as a direct metaphor. (See Steen et al., 2010, pp. 38−39 for further explanation.)

3.2.3 MZ (metaforazászló, 'metaphor flag')

When a word or morpheme potentially signals that a cross-domain mapping may be at play, it is marked as a metaphor flag. In the previous example, the morpheme -*szerű* ('like') does not prompt a cross-domain mapping in itself, but it highlights the metaphorization that the other component of the construction undergoes, directing the reader's attention to the intended metaphorical interpretation. Therefore, the morpheme functions as a metaphor flag. (Steen et al., 2010, pp. 40−41 discusses the category with additional examples.)

3.2.4 MKKimp (metaforához kapcsolódó implicit kifejezés, 'metaphor-related implicit expression')

The MKKimp tag was introduced to indicate that the annotator encountered a term that is in a coreferential relationship with a metaphorical expression and thus contributes to the maintenance of metaphorization, even though it is not the initiator of the cross-domain mapping. In the following example: *hogy őrködhessen az ottani javainkon, no meg hogyez-alatt* [this-under 'during this time'] *a bevételről se kelljen egészen lemondanunk* ('in order to take care of our possessions, and to avoid having to give up on our entire income during this time'), the highlighted expression (*ezalatt* 'during this time') refers back to the antecedent *őrködhessen* ('take care') as a pronominal anaphor. Since the expression *őrködhessen* ('take care') was labelled MKK in prior analysis because its contextual meaning ('keeping an eye on the possession of the others') is a metaphorical extension of the basic meaning ('as a guard, (s)he shall serve somewhere'), its anaphoric counterpart is identified as an instance if implicit metaphor. (For additional examples within the framework of MIPVU see Steen et al., 2010, pp. 39−40)

3.2.5 MKI (metaforához kapcsolódó inflexió, 'metaphor-related inflection')

The MKI label indicates inflections that do not express a concrete, specific, and spatial meaning (which is the basic meaning of these case markers according to the dictionary), but they refer rather to a conceptualization that can be interpreted as an extension of the basic meaning on the basis of similarity. For example, the inessive inflection has a concrete, specific, human-oriented basic meaning of 'being in(side) something'. In the following construction, however, the structure invites the conceptualization of a weather phenomenon as a container: *ellen-szél-ben* (counter-wind-ine 'in headwind'). Thus, the case marker makes a cross-domain mapping between the concept of wind and the concept of a container possible.

3.2.6 MKA (metaforához kapcsolódó argumentum, 'metaphor-related argument')

The MKA label is used to denote those arguments that contribute to the metaphorization of their verb head, but themselves are not the initiators of cross-domain mapping. In other words, these words play a role in metaphorization not in themselves but in a tight connection to a verb, i.e., as the component of a construction. For instance, in the following construction: *az érzékenység-em-et akar-ta érint-eni* (the sensibility-poss3sg-acc want-pst.3sg touch-inf '[she] wanted to refer to my sensibility') the word *érzékenységemet* ('my sensibility') is the direct object of the infinitive *érinteni* ('to touch', it is a metaphor-related expression in this construction). This argument also gives rise to an extension in the meaning of the infinitive: it specifies the type of the process symbolized by the infinitive, making an abstract contextual meaning ('referring to something') possible. Therefore, the nominal argument of the infinitive can be identified as a metaphor-related argument. Note that in accordance with our methodological decisions, the case marker of the argument (here it is the accusative marker) is not analyzed separately from the expression as a whole.

3.2.7 MKKomp (metaforához kapcsolódó komponens, 'metaphor-related component')

The MKKomp label is used to indicate those elements of the structure which do not initiate metaphorization in themselves, but they are involved in the emergence of a metaphorical meaning of a complex morphological, compound or attributive construction. Since one element of these complex structures has a metaphorical contextual meaning, this effect may also extend to other elements of the construction. As an example, in the following expression: *a kabát-om szárny-a* (the coat-poss1sg wing-poss3sg 'the tail of my coat') the nominal element *szárnya* ('its wing') has metaphorical meaning since instead of the primary meaning of the noun *szárny* ('the part of an animal or imaginary creature that can fly'), a contextual meaning ('the opening part of a piece of cloth') prevails. Thus, the nominal is labelled as a metaphor-related expression in the annotation, and the nominal possessor of the possessive construction (namely *kabátom* 'my coat') gets involved in the unfolding metaphorical meaning as a component (being linked to the metaphorical core).

3.2.8 MKKid (metaforához kapcsolódó idiomatikus kifejezés, 'metaphor-related idiomatic expression')

When a word is part of an idiomatic expression the use of which may potentially be explained by some form of cross-domain mapping regarding the nonliteral meaning of that idiom, the word is marked with this label. For example, in the idiom *rend-ben a széná-ja* (order-ine the *hay*-poss3sg 'everything is all right about him' lit.: 'his hay is in order') the nominal stem *széna* has the prior, literal meaning 'mown and dried grass or other fodder crops as animal feed'. But in this construction, it evokes circumstances, emotions, or a situation in general, thus the noun *széna* initiates here a cross-domain mapping, and it can be tagged as a metaphor-related idiomatic expression.

3.2.9 MKAid (metaforához kapcsolódó idiomatikus argumentum, 'metaphor-related idiomatic argument')

When an argument is related to a verb in an idiomatic expression, and the use of the idiom may be explained by some form of cross-domain mapping from a more basic (i.e., literal) meaning of that idiom, then the argument is marked as a metaphor-related idiomatic argument. For instance, in the construction *az ő kedv-é-t kell nézn-em* (the she mood-poss3sg-acc must look-inf-1sg 'I must please her' lit.: 'I must look at/focus on her mood') the *kedvét* nominal is both a component of the conventional idiomatic expression *kedvét nézi/keresi valakinek* 'look at/focus on somebody's mood' and an argument of the verb *néz* 'look at'. The MKAid tag can mark both functions of the nominal argument.

3.2.10 MKKompid (metaforához kapcsolódó idiomatikus komponens, 'metaphor-related idiomatic component')

When a component belongs to an idiomatic expression whose use may potentially be explained by some form of cross-domain mapping from a more basic (i.e., literal) meaning of that idiom, the component is marked as a metaphor-related idiomatic component. For example, in the construction *megfelelő keret-ek között* (appropriate frame-pl within 'within appropriate compass') the nominal head *keret* basically means 'a structure that surrounds, holds, protects an object, device or image and validates it as a unit' according to CDH, but in this expression, it has the contextual meaning of 'the limitations of possibilities and conditions', hence it can be labelled as a metaphor-related expression. The metaphorization of *keret*, however, has an effect on the other component of the construction, namely the postposition *között* as well: by default, it means 'in a place bounded by persons/things', but in the above idiomatic expression it takes part in the emergence of metaphorical meaning in the construction since it refers to an abstract region of possible action.

To summarize, the following table presents the labels of the MetaID protocol in comparison with the adopted tags of the original process.

3.3 Semantic relations in metaphorical expressions

Importantly, the MetaID annotation protocol has another novelty that paves the way for the identification of not only single metaphor-related components of a text but also complex metaphorical structures. To achieve this goal, the research group developed a complete tag set for marking semantic relations between metaphorical components. (It is called the mrel tag set.) New relational tags were based on theoretical foundations laid by Cognitive Grammar (Langacker, 1987, 2008) since the systematic description of semantic integration provided by this framework proved to be an optimal point of departure for exploring the structural organization of metaphorical meaning.

During the annotation process, the starting point of tagging semantic relations is always the metaphorical component that triggers metaphorization. In other words, these tags are not assigned to a particular morphological unit but rather they connect two such units, i.e., two components of a complex metaphorical structure which have previously been labelled as metaphorical. After selecting the primary metaphorical component, it is determined whether other components also participate in the metaphorization process. Metaphorical relations have been observed in the following six cases and described by six tags accordingly.

3.3.1 The "tr" and "lm" relations

The first two tags stand for the terms Trajector and Landmark, which mark the primary and secondary focal participants of the conceptualized event (see Langacker, 2008, pp. 70–73). Although the paper introduced them previously (see 2.2), we are providing now a more detailed description of these relations.

Consider the following two examples. *Apró rovar-ok torkoskod-nak a messzire sárgálló virág-ok nektár-já-ból* (tiny insect-pl nosh-3pl the far-away yellowing flower-pl nectar-poss3sg-ela ('tiny insects nosh from the nectar of faraway yellowing flowers'). It is suggested that since *torkoskodik* 'to nosh' is a human activity in its basic meaning, the verbal stem *torkoskod-* is a metaphorical component. But the fact that this metaphorical verb refers to the agents of the process—the trajector in cognitive grammatical terms, namely *rovarok* 'insects'—makes the non-human entity observable as a human one. Based on this relation it is argued that the primary figure of the verb (expressed by the nominal) also participates in the metaphorization process by acquiring non-inherent characteristics (i.e., being able to nosh) due to the effect of a metaphorical component together with which it constitutes an event.

In the second complex construction, the landmark relationship can be highlighted: *mindegyikük* [*számítógép*] *fel tud épít-eni kapcsolat-ot bármelyik másik-kal* (each of them [computer] up.prev can build-inf relationship-acc any other-inst 'each of them can build a relationship with any other'). By examining the process expressed in the example, *felépít* 'build' can be identified as metaphorical because it should not be understood in its basic meaning connected to physical activity. Moreover, since *felépít* 'build' would normally refer to a building, the landmark of the process, i.e. *kapcsolat* 'relationship' can be perceived as a building-like object.

3.3.2 The "ela" relation

The tag ela 'elaboration' was used partly as a consequence of morpheme-based annotation. In cases where a morpheme is metaphorical, it can trigger the metaphorization of the whole word (see also Kaya, 2019, p. 233 in relation to Uzbek). Elaboration indicates in our protocol a non-specified elaborative operation, namely that a metaphor-related morpheme affects the meaning of other components on a morphologically or syntactically complex structure extending the potentiality of metaphorization to them. (For more on semantic elaboration, see Langacker, 2013, pp. 198−205.)

The tag ela 'elaboration' was used within word forms as composite structures, whereas "tr" and "lm" tags are assigned to multi-word expressions. The scope of the elaborative relationship includes three cases: it mainly occurs within the composite structures of a metaphor-related inflection and a metaphor-related component (MKI + MKKomp), between the components of a metaphorical compound (MKK + MKKomp) and between a metaphor-related adjective and a word whose referent it describes (MKK + MKKomp again, but in a syntactic construction). These phenomena are illustrated in the following.

The first case is when the suffix indicates and carries metaphorization, as in the following example: *az el-múl-t föld-történeti kor-ok-ban* (the away.prev-pass-ptcp earth-historical period-pl-ine 'in the past geological periods'), where the spatial metaphorization of time unfolds through the inessive case marker. The suffix *-ban* 'in' is used here as a time reference; however, its basic meaning would refer to space. According to cognitive metaphor theory, *korok* 'periods' are conceptualized here as containers. Based on the idea that it is the suffix which initiates the metaphorization process spreading it to the whole word form, there is a general elaborative semantic relation between the inflection and the nominal stem. It is worth mentioning that Hungarian postpositions can also express time relations. In the former example *megfelelő keret-ek között* (appropriate framework-pl between 'within the appropriate framework'), *között* 'between' does not refer contextually to a physical place according to its basic meaning; instead, the postposition specifies an abstract state. Because of this cross-domain mapping the noun *keret* 'framework' can be also construed as a physical place, to which *között* 'between' would refer in its basic meaning. Therefore, there is a potential elaborative relationship between the postposition and the nominal stem.

Secondly, a compound word can become metaphorical because of one of its components. The nominal compound *Krizantém-öl* ('chrysanthemum-lap') from a contemporary Hungarian poem depicts the figure of Autumn as a female character focusing on one part of her body. Since the second component of the expression initiates a cross-domain mapping between the season and a female figure, it also makes the metaphoric interpretation of the first noun possible (in which the flowers form the body part of the metaphoric lady). The first noun is involved in the process of metaphoric meaning creation as a component and becomes elaborated by the second noun. Therefore, their semantic relation can be labelled as elaboration.

And finally, a metaphorical adjective can have a semantic effect on a word whose referent it characterizes. For instance, in the expression *súlyos következmény-ei is lehet-nek* (heavy consequence-poss-pl too be.possible-3pl 'it may result in heavy consequences') the noun *következmény* 'consequence' is modified by the adjective *súlyos* 'heavy'. The latter is not to be understood in its basic meaning ('heavy, weighty'), instead, it refers to the fact of seriousness. At the same time, since *súlyos* is connected syntactically to and describing the nominal stem *következmény* ('consequence'), the latter could be also construed as a physical object. Hence, it is appropriate to draw an elaborative relationship between the adjective and the nominal form.

3.3.3 The "poss" relation

In addition to the "ela" category, another label is needed to identify a specific sub-type of semantic integration within composite structures. In the Hungarian expression *a lépcső töv-é-ben* (the stairs stock-poss3sg-ine 'at the bottom of the staircase') there is a possessive relation between the staircase and one part of it (namely, the lower part), expressing a part-whole relation between them. Literally, the stock is possessed by the staircase. In addition, the noun *tő* ('stock') initiates a cross-domain mapping by which the staircase is conceptualized as a plant-like entity. The relation tag "poss" makes this connection marked, with the aim of distinguishing general elaborative relationships from structures where metaphorization is based on a part-whole relationship expressed by a possessive construction. In these expressions, it is the metaphorical meaning of the part which becomes the source of metaphorical meaning generation. In other words, it can be considered an instantiation of the more general reference-point configuration (see Langacker, 2013, pp. 81–85 for a cognitive grammatical description), which motivates again the distinction between the elaborative and the possessive relations.

3.3.4 The "expm" relation

The "expm" label is the abbreviation of 'explicating metaphorical meaning'. When there is an explicit signal of metaphorization in the discourse (metaphor flag in the terms of the MIPVU protocol), a specific semantic relationship can be identified between the signal and the expression used as a direct metaphor (MZ + dMKK in our tag set). One example of this is *"menyegző-jé-t"* (wedding-3sg-acc 'her wedding') where the square quotes indicate that the expression *menyegző* ('wedding') needs to be comprehended metaphorically in the given context (describing the reproduction process of plants in spring). Thus, the square quotes function as metaphor signals and the relationship between them and the nominal expression is explication.

3.3.5 The "r" relationship

Introducing this tag had methodological reasons: preverbs can either contribute to or modify the meaning of verbs with which they constitute a single dictionary entry with a basic meaning. However, according to the orthographical conventions of Hungarian, preverbs can be separated in several sentence structures (because of inversion or auxiliaries being inserted between the preverb and the verb). Therefore, the "r" relation tag makes it clear that the two elements belong to each other and form a metaphorical expression together. In the former sample sentence *mindegyikük* [*számítógép*] *fel tud épít-eni kapcsolat-ot bármelyik másik-kal* (each of them [computer] up.prev can build-inf relationship-acc any other-inst 'each of them [i.e., the computers] can build a relationship with any other') it can also be seen how the "r" tag works, connecting a separated preverb with its verb.

To summarize the theoretical and methodological innovations of our "morpheme-based" MIPVU, we can conclude that the MetaID protocol can be used not only for examining the structural patterns of metaphorization in an unprecedentedly detailed

manner but also for investigating the semantic complexity as well as the prefabricatedness of metaphorical expressions. In other words, it can be considered both as a major specification of MIPVU and as the beginning of a new phase in quantitative metaphor identification and the qualitative, corpus-based analysis of metaphorization. To render the application of our hybrid approach possible, the following section presents annotation guidelines in a step-by-step manner.

3.4 MetaID annotation guidelines

The guidelines for the annotation process are as follows.

(1) Annotation of metaphor-related components in Hungarian texts

(a) Divide the text into morphological units.
(b) Determine the basic meaning of the morphological unit according to CDH. Determine the contextual meaning of the same morphological unit, again relying on CDH.

  (i)   If the contextual meaning does not coincide with the first meaning specified by CDH, then decide if there is a semantic relationship based on indirect meaning generation between the first and the identified contextual meaning of the morphological unit.
  (ii) Investigate if this indirectness can be explained by metonymization or generalization, and if the contextual meaning can plausibly be supported by either of them, leave the unit untagged.
  (iii) If the indirect semantic relationship between the first and the contextual meaning seems to be based on cross-domain mapping, assign the MKK or MKI tag to the unit.
  (iv) If cross-domain mapping is reflected explicitly by a morpheme or a punctuation mark, assign the dMKK tag and the accompanying MZ tag to the components of the metaphorical structure.

(c) Identify the arguments of the verbal stem receiving MKK tag and determine if they contribute to its metaphorization as a head of the composite structure. If this is the case, assign the MKA tag to the verbal argument.

  (i)   If the annotator, relying on their professional intuition, assumes that the morphological unit initiating metaphorization forms an idiomatic expression together with its argument or the other component, check the strength of the collocation in the reference corpus. If the LogDice typicality score is higher than 8.00, assign MKKid tag to the head and MKAid tag to the argument of the verbal construction, or MKKompid

> tag to the other non-specified component of the idiomatic noun phrase. (For the technical details of measuring idiomaticity see Sect. 4)
>
> (ii) If a morphological unit is in a coreferential relationship with another morpheme proving to initiate cross-domain mapping (thus receiving MKK tag), then allocate the MKKimp tag to the coreferential pronominal or nominal expression.

(d) Also identify the noun stem of the inflection receiving an MKI tag and assign the tag MKKomp to it in all cases.

(e) If a preverb received an MKK tag, then assign an MKKomp label to the verbal stem as well.

(f) In cases when a nominal stem identified as metaphorical functions also as a possessed entity in a possessive construction, then assign the MKKomp tag to the nominal stem functioning as the possessor in the construction.

(2) Annotation of the semantic relationship between metaphorical components of Hungarian texts:

(a) If the morpheme annotated with an MKK tag is a verb stem, determine if the semantically accompanying word form annotated with an MKA tag is the elaboration of the primary or the secondary schematic figure of the verb's meaning.

> (i) If it is a primary figure, mark the connection between the verb and its argument with the "tr" relational tag.
>
> (ii) If it is a secondary figure, mark the connection between the verb and its argument with the "lm" relational tag.
>
> (iii) In the case of idiomatic constructions receiving MKKid and MKAid tags, follow the same protocol.

(b) If an inflectional morpheme is annotated with the MKI tag, mark the relationship between it and the noun stem tagged with the label MKKomp with the "ela" relational tag.

(c) If a preverb is annotated with the MKK tag, also mark the relationship between the metaphorical preverb and the verbal stem annotated as MKKomp with the "ela" relational tag.

(d) If a noun is annotated with the MKK tag because of being the target component of a metaphorical reference-point structure and the reference point is annotated with the MKKomp tag, then mark the semantic relationship between the target and the reference point with the "poss" tag.

(e) Between metaphorical components annotated by dMKK and MZ tags, mark the connection with the "expm" relational tag.

(f) When a metaphorical verb stem and its preverb are not adjacent in terms of linear order, mark the non-specified grammatical relationship with the "r" relational tag.

To provide a brief comparison between the MetaID protocol and the original method (on the basis of Steen et al., 2010, pp. 25−42), the most important departures from MIPVU can be summarized as follows. While both methods attempt to find metaphor-related expressions in a text in a systematic way, our protocol proceeds on a morpheme-by-morpheme basis instead of relying on lexical units. Despite this profound change, derivational morphemes which change the part of speech category of the original word are handled in the same way as in MIPVU. Moreover, if the case marker on the nominal stem belongs to a verb + argument construction morphosyntactically, the MetaID method does not identify it in itself as metaphorical. The second point of departure is that our protocol extends the examination to the argument structure of verbs, introducing a novel label to the arguments and contributing to metaphorization on the level of the construction. Thirdly, we strive to explore the internal organization of morphologically complex formations (inflected nouns or verbs with preverbs), highlighting components which are not the triggers of a potential metaphorical meaning but are also involved in the process of metaphorization. Our fourth innovation is to indicate the idiomatic or prefabricated character of a multi-word metaphorical expression (these are, mainly, argument constructions) on the basis of collocational behavior. Finally, a new set of relational tags has been developed in order to make the semantic organization of metaphorical expressions possible. The implementation of a morpheme-based identification of metaphorical meaning required, however, not only theoretical but also practical improvements as well. The next section gives an overview of the latter.

4 Implementation of the MetaID protocol

The following section deals with the details of how the authors implemented the adapted hybrid protocol and what preliminary results have come from the testing phase of the procedure. Both the research infrastructure and the reliability test are detailed here, even in a nutshell.

4.1 Improvements in the design of annotation

For the process of metaphor identification, it was necessary to choose an appropriate dictionary to define the basic meaning and distinguish it from the contextual meaning of the morphological units in the text. In the absence of a corpus-based dictionary of Hungarian (the situation being similar to the Lithuanian, Uzbek and Sesotho languages among others), we chose The Concise Dictionary of Hungarian (CDH) for the task of sense disambiguation, because of its partial corpus-based elaboration (word-frequency data are drawn from the Hungarian National Corpus (Oravecz et al., 2014) in this dictionary), and since this is the most adequate and complex dictionary for Hungarian in our time. For mapping the verbal constructions during the annotation process, the Dictionary of Hungarian Verb Structures (Sass et al., 2011) was used by the annotators. Morphological analysis and segmentation of lexemes

were carried out by the morphological analyzer of the e-magyar Digital Language Processing System (Váradi et al., 2018). Using an NLP tool instead of relying on the professional intuition of the annotators enhances the reliability of the otherwise more fine-grained analysis.

As mentioned above, a new subprocess of measuring the idiomaticity of the expressions and their argument structures was developed and integrated into the protocol. For this specific purpose, a computational measuring tool, namely the word sketch browser of the Hungarian Web 2012 corpus (huTenTen12) and its collocational values were exploited. On the one hand, this database made our protocol more transparent and therefore more reliable; on the other, it sped up the whole procedure. Decisions on labelling idiomaticity were based on the logDice typicality score (Rychlý, 2008), which—after the selection of possible candidates for idiomatic expressions on the basis of the annotator's intuition—indicates how strong the current collocational relation is. (For the advantages of choosing logDice over other scores see Gablasova et al., 2017). The higher the score is (above 8.00 in our study), the stronger is the association between the node and the candidates of a collocation. In cases surpassing our predetermined threshold, we used the MKKid tag for annotating metaphor-related idiomatic expressions and the MKKaid tag for indicating their argument structure or MKKompid (see Sects. 3.2.8−3.2.10).

As an online collaborative process, the annotation was implemented as a trial on the WebAnno surface designed by CLARIN Research Infrastructure for Language Resources and Technology (Castilho et al., 2016). This platform makes the parallel work of more than one annotator possible, providing appropriate visualization for both tagging the elements and their semantic relations in the texts. Moreover, it can be used for sub-token (i.e. character-based) annotation as well. An additional advantage of the tool is that it allows annotators to work on the same text (with the same tag set) individually and to compare the individual annotations on a more general level. In the course of annotation, annotators cannot see each other's works. After each annotator finishes the individual work on the corpus, the inter-annotator scores (hence, the level of reliability of the process) are automatically calculated by the platform and then, during curation, individual annotations can be merged into a single discussed version.

Figure 1 presents the online collaboration process via a highlighted segment of the annotated text. Initially, the software splits the whole text into sentences. The labels of the metaphorical elements are shown above the chosen morphological units; meanwhile, the relations between components are indicated with arrows and relational tags between them. With the use of WebAnno, the process of the annotation can become more perspicuous, and the work of annotators can be easily tracked by the project manager in contrast with offline annotation. Finally, the results of individual annotation processes can be compared, revised and unified in a curated document, which serves to display the integrated results of the collaborative procedure.

With the help of these practical decisions, the MetaID protocol maintains the dictionary-based methodology of metaphor identification, adjusting it nonetheless to the opportunities inherent in contemporary computer-assisted research and making the original method open to semi-automatized solutions. This is the second reason for defining our protocol as a hybrid process of metaphor annotation.
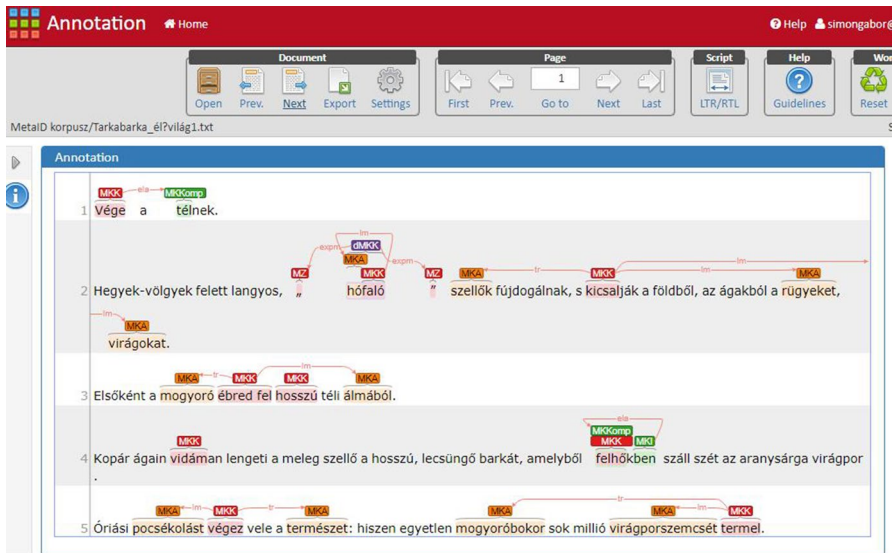
Fig. 1 Online annotation on the WebAnno surface

4.2 The preliminary results of testing the proposed protocol

In the process of developing and testing our hybrid annotation system, we started to work with two short encyclopedia entries (32 and 58 words), and a novel excerpt (324 words). After the first phase of annotation, the test corpus was expanded with an excerpt of an informational prose text for children (498 words, enlarging the whole corpus to 908 words in total) divided into five parts for pairwise annotation. To create a manageable and accurate procedure of analysis, it was obvious that simple texts were needed at the beginning, without many creative and/or non-conventional metaphorical expressions. This way of organizing the process of annotation had the advantage of measuring inter-annotator agreement and hence testing the reliability of the annotation schema.

As a second step of testing the adapted method's reliability and also working on more concrete problems related to the metaphors, we applied the method to a research corpus of a larger scale. The material of the small-scale annotation was kept and expanded to a reference corpus with excerpts from various text types (abstracts of research papers, informational prose for children, comments from online discourse, news, reports, and fiction), adding up to a total of 1940 tokens in this subcorpus. The reference corpus served as a basis for a comparison between genres presented in the other two subcorpora, namely literary and informational texts. (For a similar study see Krennmayr 2015). These two subcorpora reached the approximately equal size of 1997 and 2039 tokens. The structure of the material was planned to be well-balanced, containing four sections of approximately 500 tokens in each subcorpus. The four literary genres were lyric poetry, epic prose, drama texts and lyrics of popular songs. The subcorpora of informational text types contain an

equal amount of interview texts, informational prose texts, dictionary entries and an excerpt of a research paper. Consequently, the total size of our corpus amounted to 5976 tokens, and 8 professional annotators worked on metaphor identification in the corpus.

The elaboration of our hybrid method as well as the extension of the tag set was carried out in two phases in our research. As has been demonstrated in the former section, the first phase was a small-scale annotation process, in which one part of the test corpus (47.2%) was annotated by each member of the group and the other (52.8% of the whole corpus) was divided into subparts and tagged by pairs, i.e., by two independent annotators. (Altogether 6 annotators worked in this phase of the project.) This combined process made it possible for the whole group to get enough practice in the original protocol and its adaptation; additionally, we were able to carry out an in-depth analysis of a smaller sample of the corpus.

Since most of the texts were processed in pairs, the appropriate coefficient for assessing the reliability of the schema is Cohen's kappa (Artstein & Poesio, 2008, pp. 559−560, 561−562). Degrees of the inter-annotator agreement were calculated by the WebAnno platform. The means of the kappa-values were 0.928 by mtags and 0.923 by mrel. Since the global performance of the annotators is above the 0.8 threshold in kappa statistics (see Carletta 1996: 252, Artstein−Poesio 2008: 22), even the first version of the adapted schema can be considered reliable.

During the extension of the test corpus to a specific research corpus, however, new theoretical problems and language-specific difficulties arose. Taking samples from different text types into the corpus confronted us with previously unknown structural phenomena of metaphorical expressions in Hungarian. Thus, additional categories had to be added to our tag set and some further specification of the identification process was required. The refined set of labels and the enlarged test corpus made it necessary to measure the reliability of the process again. In other words, it was not the performance of the group that led us to reassess inter-annotator agreement but the need for an adjustment in methodological adaptation.

In the course of the second, large-scale phase of the annotation, each member of the group (7 annotators in total) worked as an independent annotator of the entire corpus. For this reason, we changed the coefficient from Cohen's kappa to Krippendorff's alpha. The latter has several advantages in the case of multiple coders: it provides a suitable method for differentiating between disagreements, moreover, it calculates with the totality of judgements instead of using individual differences (for a description of the mathematical formulas of the coefficient, see Artstein & Poesio, 2008, pp. 564−567, Artstein, 2017, pp. 301−302).

The alpha coefficient has a stricter threshold of reliability: $\alpha = 0.8$ is considered the lowest level of acceptance (Artstein & Poesio, 2008, pp. 576). Against this benchmark, the second version of the adapted MIPVU schema proved to be reliable too: in the large-scale annotation, the overall alpha-measure was exactly 0.80 by mtags and 0.82 by mrel in the whole corpus. (The coefficient was calculated by the WebAnno surface again.) The details of the results are demonstrated in Fig. 2.

First of all, it is worth noting that out of the six measurements only three surpassed the consensual limit of acceptance. Informational and scientific texts set the annotators a complex task, which is in accordance with the proportions of
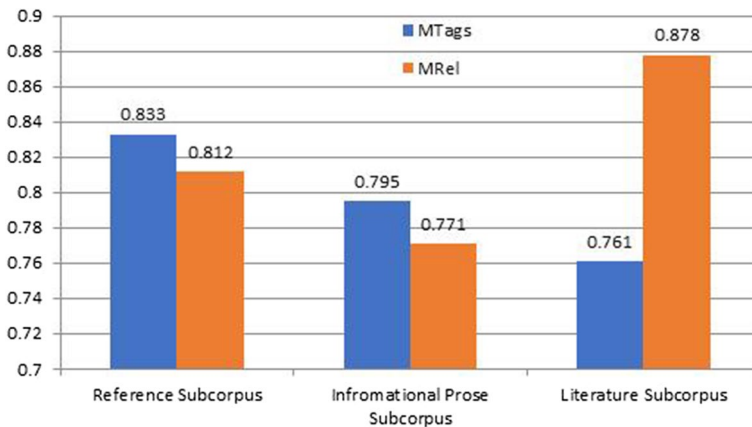
Fig. 2 The inter-annotator agreement in Krippendorff's alpha

metaphorical data in these texts. The overall reliability of the process is acceptable, nevertheless, the results clearly demonstrate that additional refinement of the schema is required for better performance.

There are differences both between tag sets and subcorpora. The general performance of the annotators in using the mtags set is poorer in the entire corpus despite the overwhelming agreement we find in the realm of metaphorical structures and components in two of the three subcorpora. The best agreement produced in the Literature subcorpus in labelling metaphorical relations makes the latter a more reliable subschema. Put simply, identifying metaphorical expressions in the corpus seems to be a less clear task than tagging semantic relations between their components. However, the efficiency of the schema is different regarding the particular text types: while the reference corpus (with its mixed sampling) is the most unambiguous in both of the subsets, the annotation of metaphorical components in literary texts is a more complicated task in contrast to labelling semantic relations in the same subcorpus. Consequently, both the reliability of the hybrid MetaID protocol and the language-specific and/or topic-specific challenges of metaphor identification can be explored by measuring inter-annotation agreement.

5 Conclusions and future perspectives

The main goal of the paper was to present the first systematic attempt to apply MIPVU to Hungarian and extend its scope to cover semantic relations and idiomaticity of metaphorical expressions, thereby bringing Hungarian onto the international map of the method, as well as bringing the method closer to corpus-driven pattern analysis. The results of the adaptation can be summarized as follows. On the one hand, a specific hybrid method for metaphor identification was proposed and detailed with regard to the theoretical as well as practical innovations. On the other hand, it extended the scope of the original protocol not only with language-specific

solutions but also by additional specifications of morphological analysis, by measuring idiomaticity and exploring semantic relations in order to characterize the linguistic manifestation of metaphors in a more fine-grained manner.

The main challenge that we faced was the need to modify the original method to handle the agglutinative character of Hungarian. In response to this challenge, we changed the level of annotation from lexical to morphological units for detecting metaphorization at a sub-token level. This morpheme-based method required that we (i) define some principles for localizing the units of the analysis before tagging, and (ii) introduce new tags which refer to inflectional metaphorization (MKI), verb + argument constructions (MKA), and other components contributing to metaphorical meaning (MKKomp). The adapted and extended MetaID tag set allowed us to identify metaphorical expressions that are specific to a language with complex morphology like Hungarian by analyzing the components of compounds and dealing with preverbs and inflectional morphemes as metaphor-related components. In this respect, our "morpheme-based" version of MIPVU (the MetaID protocol) can serve as a point of departure for novel adaptations to agglutinative languages (both within and beyond the Finno–Ugric language family).

Since using morphemes as alternative basic units of analysis is a radical change, some may have doubts about whether prospective results (gained by our new protocol) can still be compared with data extracted by the original method. One concern can be that our decision considerably increases the number of annotated elements in the corpus, which in turn makes the overall proportion of metaphorical elements appear less similar to the picture obtained by MIPVU. For example, metaphorization can be assigned to linguistic elements on the token level and on the sub-token level as well in our method, depending on the actual morphological complexity of the expression. However, morphemes which are able to behave metaphorically in Hungarian (e.g., preverbs or case markers) do have their equivalents in English as well: prepositions are marked in MIPVU either as MRW or as WIDLII elements. Thus, although the distribution of metaphorization becomes more subtle in the analysis based on the MetaID protocol, the final number of tagged elements does not need to be remarkably different.[6] In other words, a morpheme-based annotation protocol does not necessarily make a language appear "more metaphorical", but it does allow for a careful detection of the complexity of metaphorization. Another problem of comparability may derive from the more diversified tag set proposed by the MetaID protocol: due to the introduction of the inflection-related label and further tags for marking semantic relations (see the extended tag set later), our protocol highlights the grammatical organization of metaphors in a way that is hidden by the MIPVU method. But this shift of perspective does not imply a new concept of metaphor in language: it is still defined as a form of indirect meaning generation motivated by

---

[6] Moreover, in the case of other morphologically rich languages it is the whole lexical unit which receives a tag even if only one of its components is responsible for metaphorization. In other words, a more or less hidden morpheme-level annotation does exist in other adaptations of the MIPVU with no modification in the tagging conventions. Thus, the numbers emerging from these alternative protocols and from our method remains comparable regarding their order of magnitude.

cross-domain mappings. To put it differently, the MetaID protocol does not consider metaphor a morphological phenomenon, but rather allows for the exploration of the internal structure of metaphorical expressions in a more delicate manner. Since our method is based on the original protocol, it is an extension of that towards a more comprehensive process of metaphor identification that adds a new layer to linguistic analysis.

At this point, it is worth mentioning too that although the inflectional label of metaphorization does not necessarily increase the total amount of metaphor-related expressions in the corpus, the same cannot be said about the argument-tag and the component-tag. Since these labels indicate the contribution of a structural component to a head element triggering metaphorization, in these cases the phenomenon is identified in two ways. In order to avoid redundancy, the annotator can calculate with a verb + argument construction as a broader unit (this way the additional tags do not increase the total number of metaphors). But the specific types of metaphor-related components can also be taken into consideration, making a more nuanced analysis possible.

Another theoretical challenge was to develop the original method with the aim of precisely accounting for the complex ways in which metaphors are linguistically manifested. To achieve this, (i) we made a more rigid distinction between metaphor, metonymy and generalization, and (ii) developed a method for recognizing and analyzing idiomatic expressions on the basis of their collocational behavior. Prefabricated expressions were handled in our protocol with the tags of MKKid, MKAid and MKKompid. In this way, we were able to detect potential cases of metaphorization consistently and make the annotation process as transparent as possible, hence a greater level of accuracy was achieved in metaphor identification.

Finally, the MetaID protocol includes a new tag set for exploring and annotating semantic relations between the components of a metaphorical expression. Based on the systematic description of the process of semantic integration defined by the framework of Cognitive Grammar, our relational tags—the trajector and landmark, the possessive, the explicative and the non-specified relation tag—can reveal the semantic complexity of metaphorical expressions, besides their structural organization.

Beyond these theoretical issues, the operationalization of the original methodology posed some challenges as well. Without any comprehensively corpus-based dictionary of Hungarian, The Concise Dictionary of Hungarian (CDH) proved to be the most appropriate choice for defining the meanings of morphological units in the texts as a baseline for analyzing their partial corpus-based elaborations. Additional databases were also applied for the identification of metaphorical verb + argument patterns and for calculating the collocability of components in particular expressions. In line with the innovation of morpheme-sensitivity, an automatized morphological analysis was integrated into the protocol by using an NLP tool for Hungarian. The trial process of collaborative annotation (performed by 7 annotators) was implemented on the WebAnno surface, which also makes it possible to measure the scores of inter-annotator agreement in the future.

After the successful adaptation of MIPVU to Hungarian and the elaboration of its hybrid design, it is the careful analysis of its reliability which is very much on the agenda of the research. Although the first test phase of implementing the morpheme-based MetaID protocol is a promising start, this does not mean that minor changes are not required. We need additional large-scale annotation and examination for further evidence of the productivity of our proposal. Since the empirical investigation is in progress now, its results are due to be reported in the near future.

A reliable method of metaphor identification in Hungarian makes further qualitative and cross-linguistic investigations possible as well. First of all, it gives a solid ground for an in-depth study of verb-centered metaphorical expressions in terms of their construction-like organization, extending the scope of construction grammars to figurative language use as well. On the basis of verb + argument metaphorical data, schematic patterns can be abstracted on different levels (e.g., the collocational structure of a particular verb + particular arguments, or more schematic patterns of a particular verb + abstract types of arguments). Then, an emerging dataset of relatively typical constructions can support the formulation of queries in other corpora, producing further evidence about the constructionalization of metaphorical meaning.

Our future perspectives include a comparative study of metaphorization in other Finno-Ugric languages. The adaptation of the method to other languages with rich inflectional morphology and a common historical origin like Finnish requires further collaboration, but our analysis can serve as the first step in that direction. In this context, the exploration of the complexity of metaphorical structures, for example by examining semantic relations between the components of a metaphorical expression or the constructions of verbal metaphors makes a significant contribution to the general description of the grammatical system of a language, and it can shed new light on cross-linguistic or typological similarities.

Moreover, measuring the idiomatic character of metaphorical enhances to shed new light on the overlap between metaphorization and prefabricatedness. This can make the meaning description in the field of lexicography more accurate, and the lexical semantic analysis of polysemy can also be refined on the basis of our methodological proposal.

The expanded annotation method and development of a manually annotated metaphor corpus have also opened up pathways for adopting new approaches from computational linguistics. As noted by Steen, "[t]he next step is for MIPVU to be automated. It would be such a help if we can run bulk analyses of large data sets in the same way across many domains between different languages" (Nacey et al., 2019, p. 13). Combining the metaphor annotation tool with NLP devices such as WebAnno and e-magyar represents a key step on this path. To conclude, we hope that the MetaID protocol paves the way for further development in metaphor identification, and the status of Hungarian as a *no man's land* in terms of MIPVU will fundamentally change even in our days.

Declarations

Competing interest The authors have no relevant financial or non-financial interests to disclose.

References

Artstein, R. (2017). Inter-annotator agreement. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic annotation.* Springer.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics,34*(4), 555–596. https://doi.org/10.1162/coli.07-034-R2

Babarczy, A., Bencze M., I., Fekete, I., Simon, E. (2010). The Automatic Identification of Conceptual Metaphors in Hungarian Texts: A Corpus-Based Analysis. In Bel, N., Daille, B., Vasiljevs, A. (Eds.), *Methods for the automatic acquisition of Language Resources and their evaluation methods.* Proceedings of the LREC 2010 Workshop (pp. 30−36).

Bogetić, K., Broćić, A., & Rasulić, K. (2019). Linguistic metaphor identification in Serbian. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), *Metaphor identification in multiple languages: MIPVU around the world.* John Benjamins.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistics. *Computational Linguistics, 22*(2), 249–254.

Castilho, R. E., Mújdricza-Maydt, É., Yimam, S. M., Hartmann, S., Gurevich, I., Frank, A., & Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. *Proceedings of the workshop on language technology resources and tools for digital humanities (LT4DH)* pp. 76–84, The COLING 2016 Organizing Committee. https://www.aclweb.org/anthology/W16-4011.

Deignan, A. (2005). *Metaphor in Corpus Linguistics*. John Benjamins. https://doi.org/10.1075/celcr.6.

Dorst, A. G., Mulder, G., & Steen, G. J. (2011). Recognition of personifications in fiction by non-expert readers. *Metaphor and the Social World,1*(2), 174–201. https://doi.org/10.1075/msw.1.2.04dor

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based learning research: Identifying, comparing, and interpreting the evidence. *Language Learning,67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. *7th International corpus linguistics conference (CL 2013)* pp. 125–127. University centre for computer corpus research on language (UCREL). https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf.

Kaya, S. G. (2019). Linguistic metaphor identification in Uzbek. In S. Nacey, A. G. Dorst, T. Krenn-mayr, & W. G. Reijnierse (Eds.), *Metaphor Identification in multiple languages: MIPVU around the world*. John Benjamins.

Krennmayr, T. (2015). What corpus linguistics can tell us about metaphor use in newspaper texts. *Journalism Studies,16*(4), 530–546. https://doi.org/10.1080/1461670X.2014.937155

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Vol 1. Theoretical Prerequisites*. Stanford University Press.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.

Langacker, R. W. (2013). *Essentials of cognitive grammar*. Oxford University Press.

Lu, X., & Wang, B. P. Y. (2017). Towards a metaphor-annotated corpus of Mandarin Chinese. *Language Resources & Evaluation,51*, 663–694. https://doi.org/10.1007/s10579-017-9392-9

Marhula, J., & Rosiński, M. (2019). Linguistic metaphor identification in Polish. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), *Metaphor Identification in multiple languages: MIPVU around the world*. John Benjamins.

Meurers, W. D., & Müller, S. (2009). Corpora and syntax. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An international handbook*. Walter de Gruyter.

Moon, R. (1998). *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Clarendon Press.

Nacey, S., Dorst, A. G., Krennmayr, T., & Reijnierse, W. G. (2019). *Metaphor Identification in multiple languages: MIPVU around the world*. John Benjamins. https://doi.org/10.1075/celcr.22.

Oravecz, C., Váradi, T., & Sass, B. (2014). The Hungarian Gigaword Corpus. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. pp. 1719–1723. European Language Resources Association (ERLA). https://www.aclweb.org/anthology/L14-1536/.

Pérez-Sobrino, P. (2014). Book Review: A Method for Linguistic Metaphor Identification: From MIP to MIPVU. *Metaphor and the Social World, 4*(1), 138–146. https://doi.org/10.1075/msw.4.1.07per

Pusztai, F. (Ed. in chief). (2003). Magyar értelmező kéziszótár. [The Concise Dictionary of Hungarian.] Akadémiai Kiadó.

Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing* (pp. 6–9). Masaryk University.

Sass, B. (2022). Principles of corpus querying: A discussion note. *Acta Linguistica Academica,69*, 599–614. https://doi.org/10.1556/2062.2022.00581

Sass, B., Váradi, T., Pajzs, J., & Kiss, M. (2011). *Magyar igei szerkezetek. A leggyakoribb vonzatok és szókapcsolatok tára* [Hungarian Verbal Structures: The Dictionary of the Most Frequent arguments and expressions.] Tinta Könyvkiadó.

Seepheephe, N. R., Ekanjume-Ilongo, B., & Thuube, M. R. (2019). Linguistic metaphor identification in Sesotho. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), *Metaphor Identification in multiple languages: MIPVU around the world* (pp. 267–287). John Benjamins.

Simó, J. (2011). Metaphors of blood in American English and Hungarian: A cross-linguistic corpus-based investigation. *Journal of Pragmatics,43*(12), 2897–2910. https://doi.org/10.1016/j.pragma.2011.05.004

Steen, Gerard J. (2007). *Finding metaphor in grammar and usage. A methodological analysis of theory and research*. John Benjamins.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). A method for linguistic metaphor identification: From MIP to MIPVU. *John Benjamins*. https://doi.org/10.1075/celcr.14

Tóth-Czifra, E. (2014). Feeling the taste of victory: The figurative conceptualization of mouth and tongue in English, German and Hungarian. In F. Polzenhagen, Z. Kövecses, S. Vogelbacher, & S. Kleinke (Eds.), *Cognitive explorations into Metaphor and Metonymy* (pp. 51–65). Peter Lang.

Urbonaitė, J., Šeškauskienė, I., & Cibulskienė, J. (2019). Linguistic metaphor identification in Lithuanian. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), *Metaphor identification in multiple languages: MIPVU around the world* (pp. 159–181). John Benjamins. https://doi.org/10.1075/celcr.22.08urb

Váradi, T., Simon, E., Sass, B., Mittelholcz, I., Novák, A., Indig, B., Farkas, R., & Vincze, V. (2018). E-magyar: A Digital Language Processing System. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation LREC 2018* pp. 1307–1312). European language resources and evaluation (ERLA). https://www.aclweb.org/anthology/L18-1208.

Veale, T., Shutova, E., & Klebanov, B. B. (2016). *Metaphor: A computational perspective*. Morgan & Claypool Publishers.