JEZIKOSLOVLJE
25.2 (2024): 269-296

UDC 811.163.42'366=111
Original scientific paper
Accepted for publication on 22. 11. 2024.
https://doi.org/10.29162/jez.2024.10

Ida Raffaelli
Matea Filko
Daniela Katunar
Faculty of Humanities and Social Sciences
University of Zagreb

Explorations of Morphotactic and Morphosemantic Transparency in Croatian – the Case of Suffixation

In this paper, we present the first study on morphotactic and morphosemantic transparency in Croatian. The study was conducted via the Survey Monkey platform and consisted of a series of tasks in which participants were required to identify morphemes or segment words with differing levels of morphotactic (MT) and morphosemantic (MS) transparency, and to provide judgments on the relatedness of words based on morphological or semantic criteria. All the motivated words from the study were produced by suffixation and belong to the general lexicon of Croatian. 101 adult native speakers of Croatian participated in the study. The study was designed to investigate the following research questions: RQ1: are suffixes and base words more accurately identified in MT transparent words, RQ2: what role morphological relations play in assessing paradigmatic relatedness between words, and RQ3: are speakers able to distinguish between derived and non-derived words. Results of the study show that both MT and MS transparency are relevant notions that need to be taken into account for a comprehensive investigation of morphological processing of Croatian derivatives.

Keywords: morphotactic transparency, morphosemantic transparency, MS patterns model, suffixation, Croatian

1. Introduction[1]

Morphotactic (MT) and morphosemantic (MS) transparency were not typically at the forefront of investigations of paradigmatic relations structuring the lexicon, the role usually taken by synonymy, semantic fields and other semantic relations. The notions of MT and MS transparency were primarily introduced by Dressler (1985, 2005) within the framework of natural morphology. As to our knowledge, avenues of research on MT and MS transparency further investigated it from: a) a theoretical perspective; b) a computational-linguistic perspective, and c) a psycholinguistic perspective. As we will discuss in Section 2, each avenue has its own goals and limitations when observing the role of MT and MS transparency in the general lexicon. This study[2] aims to expand available data on these notions with data from Croatian, a morphologically rich Slavic language that is under-researched with respect to MT and MS transparency.

The morphological structure of Croatian lexemes consists of the root or the lexical morpheme, which is the only obligatory part of the morphological structure, and one or more affixes. Affixes can be prefixes, interfixes (for compounds only, e.g., *grom-**o**-bran-Ø* 'lightning rod' < *grom* 'thunder' + *-o-* + *bran(iti)* 'to defend'), and derivational and inflectional suffixes. Each of these types of morphemes can occur more than once in the morphological structure of a single lexeme, apart from an inflectional suffix, which is obligatory in the morphological structure of inflectional lexemes and can occur only once per lexeme. Nouns and adjectives are predominantly formed via suffixation (e.g., *rad* 'work' + *-nik* > *radnik* 'worker'), and they can take up to 7 different suffixes in their morphological structure (6 derivational + 1 inflectional farthest from the root, cf. Marković 2013: 51), while verbs are predominantly formed via prefixation, and can (extremely rarely) take up to 4 prefixes (cf. Šojat et al. 2013: 92). On a general level, derivation, and more specifically suffixation, is the most productive word-formation process (Marković 2012: 124). Therefore, this study focuses on the speaker's ability to identify base words and their suffixes.

There are many morphonological changes at the morpheme / word-formation[3] boundaries in the morphological structure of Croatian lexemes. For example, the verb *predočiti* 'to envisage, to represent' is derived from the unmotivated noun *oko* 'eye' with the sound change, i.e., palatalization *k* > *č* in front of the vowel *i*:

[1] The authors thank the anonymous reviewers for their insightful comments and suggestions, which helped to improve this paper.

[2] This study was approved by the Ethics Committee of the Faculty of Humanities and Social Sciences, University of Zagreb.

[3] The word-formation boundary always coincides with the morpheme boundary.

pred- +ok(o) + -iti[4] > *predo**k**iti* > *predo**č**iti*

The verb *predočiti* 'to envisage, to represent' further serves as the base word for the derivation of the noun *predodžba* 'representation', with another sound change at the word-formation boundary (voice assimilation: voiceless *č* > voiced *dž* in front of the voiced *b*):

*predoč(iti) + -ba > predo**č**ba > predo**dž**ba.*

MT transparency is closely related to morphonological changes and refers to the (complexity of) identification of morphemic components of derived words (e.g., *lavica* 'lioness' < *lav* 'lion', without any sound changes at the word-formation boundary (*lav-ica* < *lav*), is morphotactically more transparent than *narudžba* 'order' < *naručiti* 'to order', which has a sound change *č* > *dž* at the word-formation boundary (*narudž-ba* < *naruč-iti*)[5] (see Talamo et al. 2016)). MS transparency refers to the identification of a semantic connection between the base word and the derivative,

[4] This word-formation suffix is in fact a cluster of the derivational suffix *-i* and the inflectional suffix *-ti*. Following Filko et al. 2020 & Filko 2020, we distinguish between morphological and word-formation analysis. Put simply, word-formation analysis looks for the last added affix (in our case suffix) and distinguishes it from the base word (e. g. *staklenčica* 'jar, diminutive' < *staklenka* 'jar' + *-ica*), while morphological analysis segments the entire word into its constituent morphemes. Morphological analysis can further refer to the 1) segmentation of the surface form of the lexeme into morphs (morph analysis), and 2) connection of the allomorphs to their respective morpheme (which takes the form of the single representative morph) at the deep layer (morpheme analysis). For example, the complete morphological analysis of the noun *staklenčica* 'jar, diminutive' consists of the following two layers:
1) surface layer – morph analysis: *stakl-en-č-ic-a*
2) deep layer – morpheme analysis: *stakl-en-k-ic-a.*
Morphological analysis thus enables insights into the complete morphological composition of lexemes. Word-formation analysis, on the other hand, establishes the base word in the word-formation process, as well as the stem and the word-formation affixes. The base word is the word from which the new word is being derived (or composed in case of compounds), and the stem is the part of the base word which the word-formation affix is attached to (for example, inflectional suffix has to be removed from the base word before the word-formation affixes can be added). Thus, this would be the word-formation analysis of the noun *staklenčica*:
staklenk(a) + -ica > *staklenčica*, where *staklenka* = base word, *staklenk* = stem, *ica* = word-formation suffix, a cluster of the derivational suffix *-ic* and the inflectional suffix *-a*. Note that the same derivational suffix always combines with the same inflectional suffix, and they are considered a single unit in the word-formation theory (cf. Babić 2002: 41–43).
Staklenka 'jar' is in turn derived from the adjective *staklen* 'glass.ADJ': *staklen + -ka* > *staklenka*, and *staklen* is derived from the noun *staklo* 'glass.N': *stakl(o) + -en* > *staklen*. *Staklo* is formed directly from the root *stakl.* Through these derivational steps we have obtained the complete derivational path of the noun *staklenčica* from the root to the final lexeme.
Word-formation analysis is less abstract than morphological analysis and thus closer to the average native speaker of Croatian. This is why we have used base words and word-formation affixes as the basis for our questionnaires.

[5] Note that the word *naručiti* 'to order' is itself derived from the word *ruka* with the sound change *k* > *č* at the word-formation boundary as well (*na-ruč-iti* < *ruk-a*).

i.e., the realization of a common meaning of a stem and affixes and their role in the formation of the meaning of morphologically complex words (e.g., *kućica* 'house (diminutive)' < *kuća* 'house' is morphosemantically more transparent than *maslina* 'olive' < *maslo* 'butter oil' (also see Dressler 2005)).

Alongside our motivation to expand studies on MT and MS transparency with Croatian data, our own teaching experience motivated us to develop this study. Over many years, iz showed that, although morphonological changes in Croatian are in principle regular, their identification often poses a challenge to native speakers, more so when it comes to affix identification than to base word identification. Therefore, this study investigates whether speakers can identify the components of suffixed words, i.e., base words and suffixes. For this purpose, an online study was conducted using the Survey Monkey platform.[6] Specifically, the study was designed to investigate the following research questions:

RQ1: Are suffixes and base words more accurately identified in MT and MS transparent words?

RQ2: What role do morphological relations play in assessing paradigmatic relatedness between words:
a) in contrast to semantic relations, such as synonymy, antonymy, hyperonymy, etc.?
b) in rating relations between MT and MS (non-)transparent words?

RQ3: Are speakers able to distinguish between derived and non-derived words?

The paper is structured as follows: in Section 2, we describe previous research on MT and MS transparency and point out differences with respect to the current study. We define MT and MS transparency as used in our research based on Croatian language data. In Section 3, the setup of the study is described, followed by the results presented in Section 4. Section 5 is dedicated to the discussion of the obtained results and Section 6 to concluding remarks and future work.

2. Related work and theoretical framework

At the onset of this section, it is important to note that our study focuses only on the last step of the word-formation process, i.e., adding a suffix to the base word, which can already be morphologically complex (see footnote 4). This means that our study falls outside of the scope of morphological studies dealing with affix ordering (cf. Plag 2002; Hay 2002; Hay & Plag 2004; Manova 2015; Filko 2020).

[6] The online questionnaire format was chosen, in part, to gain access to as many participants available during the COVID-19 pandemic.

Also, our study falls outside of the scope of psycholinguistic priming studies dealing with lexical decision tasks (e.g., Feldman & Stolz 1995; Feldman et al. 2002; Diependaele et al. 2009; Libben & Weber 2014, among others), which do not investigate the ability of speakers to identify and single out base words and affixes of derived words. In this respect, our study is more in line with studies of morphological awareness (Carlisle 2000; Apel 2014), i.e., the ability of speakers to understand morphological relations between words and manipulate their morphological structure. However, morphological awareness studies of English are often framed as relevant to the development of reading and writing abilities (cf. Mahony 2000; Carlisle 2000; Apel 2014), whereas Croatian as an orthographically transparent language presents fewer challenges in this regard. Moreover, our study does not test words in syntactic contexts (see e.g., Carlisle 2000) because our goal was to eliminate syntactic factors which could potentially facilitate the identification of base words and affixes of derived words. Finally, in this study, we did not examine degrees of MT and MS transparency, unlike the computational-linguistic approach to MT and MS transparency which focused primarily on the manual annotation of the degree of MT and MS transparency in derived and compound words in computational lexica (Talamo et al. 2016). Although we were not interested in this line of MT and MS transparency research at this stage, applying the notions of MT and MS transparency to a wide range of Italian lexemes, as described in Talamo et al. (2016), required a more thorough elaboration of these notions, which proved very useful for our own research.

Thus, the following sections will focus on the two notions, MT and MS transparency, and how they were adapted and interpreted for the purposes of our study.

2.1. MT transparency

Dressler (2005) discusses MT transparency in the framework of natural morphology. He claims that "the most natural forms are those where there is no opacifying obstruction to ease of perception" (Dressler 2005: 272). Purely phonological processes are on the highest end of the MT transparency scale, followed by morphonological changes, allomorphy, and weak and strong suppletion at the lowest end of the MT transparency scale. Dressler (2005: 273) also points out that word-formation bases, i.e., stems, are MT and MS most transparent when they occur in non-inflected autonomous forms, while bases which are smaller than autonomous forms are more opaque, and they can be found in Slavic languages (cf. Croatian verbal stem *četi*: *po-četi* 'to start', *za-četi* 'to conceive', *na-četi* 'to open; bite into, broach' and Dressler's English example: *re-ceive, per-ceive, con-ceive*).

Based on Dressler's insights, Talamo et al. (2016) have developed the universal scale of MT transparency for Italian in order to incorporate the information on MT transparency in DerIvaTario, a derivational lexicon for Italian. Their fine-grained

scale has 8 degrees based on the type of morphonological relation between base and derivative. The first degree is the most morphotactically transparent and is exemplified by mere juxtapositions in word-formation processes (e.g., *de-* + *tassare* > *detassare* 'to detax', Cro[7]: *vid* 'sight' + *-ovit* > *vidovit* 'clairvoyant'), while the eighth degree is the least transparent and is represented by the cases of strong suppletion (e.g., *guerra* 'war' + *-ico* > *bellico* 'war.ADJ'; Cro: *ući* 'to enter.PERF' + *-iti* > *ulaziti* 'to enter.IMPERF'). Degrees II – VII differ in the nature and the degree of MT transparency; degrees II – IV encompass different phonological and morphonological changes (e.g., *assunt-* 'to recruit, past participle' + *-ione* > *assun[ts]ione* 'recruitment'; Cro: *pjesnik* 'poet' + *-stvo* > *pjesništvo* 'poetry'), degrees V and VI deletion and alternation of constituents (e.g., *communismo* + *-ista* > *communista* 'communist'; Cro: *građanin* 'citizen.M' + *-ka* > *građanka* 'citizen.F'), and VII is dedicated to weak suppletion (e.g., *pioggia* 'rain' + *-ale* > *pluv-iale* 'rain.ADJ'; Cro: *izvući* 'to pull out.PERF' + *-iti* > *izvlač-iti* 'to pull out.IMPERF').

Hay & Baayen (2002) and Hay (2002) investigate the parsability of affixes according to their productivity and relative frequency. Relative frequency is defined as the ratio of base word and derivative frequency above which derived forms are likely to be parsed. This distinction is related to the dual-route model of morphological processing. Hay (2002: 529) hypothesizes that in cases of higher frequency of the base word, the derived word will be accessed via the decomposed route (e.g., *liberal – illiberal*, where *liberal* is more frequent than *illiberal*), while in cases of higher frequency of the derived word, the derived word will be accessed via the whole-word access route, i.e., it will not be analyzed into smaller components (e.g., *illegible – legible*, where *illegible* is more frequent than *legible*). Hay and Baayen (2002: 227) identified "a critical ratio of base and derived frequency above which derived words are likely to be parsed", i.e., are likely to be accessed via the decomposed route. They admit that the parsing line is determined mainly theoretically, based on the frequency and calculated productivity of single affixes. Two main issues arise concerning their findings. Firstly, frequencies of the derived and base words were calculated from corpora and dictionaries, which do not perfectly correspond to real-life language situations. However, their approach was detailed, and they did everything to minimize the influence of the shortages that arose from their dataset. Secondly, they empirically tested their findings on a single affix (*-al*) and investigated a group of 20 linguistics students. Their hypothesis is yet to be confirmed on a larger set of affixes and on the general population. Hay & Baayen (2003) extended their findings to the role of morphonological changes, which is related to the MT transparency scales mentioned above – affixes that do not force phonological changes at the morpheme boundaries facilitate parsing and words derived via these affixes will more likely be

[7] All examples for Croatian are our own examples adapted to the scale presented in Talamo et al. (2016).

accessed via the decomposed route. However, they do not empirically confirm their hypothesis on a larger group of speakers.

Šojat et al. (2012, 2013) discuss some MT properties of Croatian verbs. Their work is not focused on research of MT transparency; however, during the work on the morphological segmentation of more than 14,000 Croatian verbs, they noticed some of their morphotactic properties as well, e.g., they list possible prefixal and suffixal combinations in their morphological structure. Among other MT characteristics of Croatian verbs, they list the most frequent combinations of two verbal prefixes and all registered combinations of three and four verbal prefixes. For example, four verbal prefixes are found only in verbs *prepoizodnositi* (***pre-po-iz-od****nos-i-ti*) 'to carry away one by one over and over'[8] and *oneraspoložiti* 'to indispose' (***o-ne-ras-po****lož-i-ti*). Filko (2020) describes MT characteristics of Croatian nominal suffixes based on the morphological segmentation of more than 5,000 most frequent Croatian nouns. Her work focuses on morphotactic patterns of Croatian nouns with multiple suffixes in their morphological structure, i.e., on suffix ordering in Croatian. Although she counted single suffix frequency, she explored base + one suffix relations in a limited number of competing suffixes.[9]

When it comes to MT transparency in this paper, we assumed a different approach than the aforementioned research. Firstly, we focus exclusively on the identification of the base word and the final word-formation suffix in the suffixed word. This means that we are not interested in affix ordering in complex morphological structures (cf. Plag 2002; Hay 2002; Hay & Plag 2004; Manova 2015; Filko 2020), instead we observe the stem as a single (although in some cases morphologically complex) unit (see above). Moreover, our current research is not focused on word frequencies, instead, we chose words with frequent suffixes[10] in Croatian based on previous research presented in Šojat (2013), Filko et al. (2017) and Filko (2020). Regarding the MT transparency scale, it was impossible to include a fine-grained scale of MT transparency into an online questionnaire, so we decided to consider only two broad categories of MT transparency in this study. The first category encompasses all the examples with no morphonological change at the stem + suffix bound-

[8] A more accurate translation is impossible.

[9] Competing suffixes are suffixes with the same or closely related meaning, e.g., suffixes *-ina* and *-ost* both denote properties and are attached to the adjectival bases. Filko (2020) describes phonological, morphological and semantic principles which govern their choice when it comes to the derivation of new lexemes, e.g., *-ina* denotes physical and *-ost* denotes psychological properties (*visina* 'height' vs. *radost* 'joy').

[10] Single suffix frequency is counted as the number of types derived with this suffix; the greater number of derivatives with a suffix, the suffix is considered more productive.

ary (e.g., čaj 'tea' + *-nik* > *čaj-nik* 'teapot', *kuća*[11] 'house' + *-ica* > *kuć-ica* 'house. DIM', *maslo* 'butter oil' + *-ina* > *maslina* 'olive', *učitelj* 'teacher.M + *-ica* > *učitelj-ica* 'teacher.F'). The second category includes all the examples in which any kind of a morphonological change occurs at the stem + suffix boundary, with no explicit differences made between the various types of changes (e.g., *naru**č**iti* 'to order' + *-ba* > *naru**džb**a* 'order', *gra**h*** 'bean' + *-ak* > *gra**š**ak* 'pea', *ple**s**ti* 'to knit' + *-er* > *ple**t**er* 'interlace; braided ornament', *živ**ac*** 'nerve' + *-an* > *živ**č**an* 'nervous'). Finally, we wanted to see if adult speakers of Croatian are able to differentiate between derived and non-derived words and recognize bases and suffixes in derived words. Therefore, we conducted an online questionnaire to investigate how MT transparency affects their ability to decompose derived words in Croatian.

2.2. MS transparency

Research on MS transparency was usually conducted on compounds. The focus was mainly on psycholinguistic experiments related to the processing of MS transparent/opaque compounds (Libben 1998; Frisson et al. 2008; Libben & Weber 2014) or on the analysis of the transparency of compound constituents (Mattiello & Dressler 2018). Schwaiger et al. (2018) investigated derivatives, but their focus was on the MS transparency of derivatives formed with diminutive suffixes extracted from the corpora. In our study, we focus on derivatives formed with a wider range of productive Croatian suffixes from different semantic domains.

Dressler (2005: 271) defines MS transparency as "fully compositional meaning", but he points out that in word-formation "morphosemantic transparency can never be complete". He differentiates between transparent word-formation meaning and lexicalized word meaning (see also Babić (2002: 34) for the same distinction related to Croatian word-formation processes), where word-formation meaning is directly derived from word-formation processes, while lexicalized meaning can have a wider or broader meaning (e.g., *stol* 'table' + *-ar* (agentive, male) > *stolar* 'carpenter' (lexicalized meaning) vs. 'table producer' (word-formation meaning)). Thus, on the highest end of the MS transparency scale are fully compositional meanings, while on the lowest end of the scale are examples of lexicalization, i.e., examples with non-recognizable parts of the complex structure, or as Dressler (2005: 271) puts it: "Less lexicalization means more transparency, more lexicalization more opacity. More transparency implies more motivation of the compound[12] via its members."

[11] An inflectional suffix *-a* has to be removed before derivational suffix is added. The same is the case with an inflectional suffix *-o* in the following example (see footnote 4).

[12] This also illustrates the focus on compounds in the research on MS transparency.

Dressler (2005: 272) distinguishes four degrees of MS transparency in compounds (in accordance with Libben 1998):

a) transparency of both members of the compound (*door-bell*)

b) transparency of the head, opacity of the non-head member (*straw-berry*)

c) transparency of the non-head, opacity of the head member (*jail-bird*)

d) opacity of both members of the compound (*hum-bug*).

He concludes that compounds are morphosemantically more transparent than derivations and argues that the compound *dishwashing machine* is more transparent than the derivative *dishwash-er*. This indicates that derivatives will present a bigger challenge for decomposition, both formal and semantic, than compounds.

Talamo et al. (2016) applied this, slightly adapted, scale to Italian derivatives in their derivational lexicon DerIvaTario. They considered the most frequent meaning of the derivative and determined whether the meaning of the base and the meaning of the affix used in the word-formation process remained transparent in the derivative. It is important to note that they observe both bases and affixes as polysemous units with core and peripheral meanings.

2.2.1. The model of MS patterns

The theoretical background for interpreting MS transparency of the language data in this study differs in some aspects from previously mentioned research.

The investigation of MS transparency in this study is based on the model of MS patterns (Raffaelli 2012, 2013, 2015; Raffaelli & Kerovec 2008). The model of MS patterns is closely related to Guiraud's model of morphosemantic fields (1967), as well as to onomasiological approaches to lexicalization and word-formation (Baldinger 1984; Koch 1999; Grzeka 2002; Štekauer 2005). Additionally, it incorporates cognitive linguistics principles such as the lexicon-grammar continuum and the role of cognitive (semantic) processes in lexical organization (Langacker 1987, 2000; Geeraerts 1997). These elements differentiate our approach from others mentioned above from a theoretical standpoint.

According to Raffaelli and Kerovec (2008), who base the model of MS patterns on Guiraud's morphosemantic fields,[13] such a model could shed new light on morpho-

[13] Guiraud (1967) points out that lexical forms are connected and motivated by derivational, metaphorical, metonymic, and other linguistic processes. The coupling of the grammatical (derivation) and semantic processes (metaphor and metonymy in Guiraud's traditional perspective) results in the formation of various structures in the lexical system. In morphosemantic fields each derived form is related to the lexical basis (*etymon* in Guiraud's terms) in a different way. The lexical basis (which can be the base word (the unmotivated word), the root, and the stem, depending on the researcher's perspective) is the basis for various types of paradigmatic relations that are created between it and its derived forms.

logically rich languages like Croatian by considering morphological and semantic processes as relevant to the motivational paths of derived (motivated) words in the architecture of the lexicon. Since the concept of motivation is crucial for describing morphosemantic organization of the lexicon, the model of MS patterns could be integrated into approaches investigating paradigmatic relations in word-formation. These approaches (Štekauer 2014; Bagasheva 2020; Fradin 2020) view paradigmatic relations in word-formation as highly motivated, enabling the formation of multidimensional networks of relationship in the complex system of word-formation (Štekauer 2014: 368). Bagasheva (2020: 21–22) further argues that the concept of motivation influences our understanding of the strongly paradigmatic organization of the mental lexicon. While psycholinguistic research (Gagné et al. 2010) has primarily focused on compounds, the model of MS patterns, which focuses primarily on derivatives, could offer new insights into the motivation and transparency of derived lexemes in psycholinguistic contexts.

In line with the cognitive linguistics view of the lexicon-grammar continuum (Langacker 1987, 2000), the model of MS patterns does not define clear-cut boundaries between grammar and lexicon. In morphologically rich languages, such as Croatian, lexical structures are motivated both grammatically and semantically. The model of MS patterns puts emphasis on both derivational and semantic processes, considering them equally important in the formation of words and the overall paradigmatic organization of the lexicon. Furthermore, according to the aforementioned approaches to onomasiological perspectives in lexicalization and word-formation, the model of MS patterns accounts for diverse derivational and semantic processes as active processes, i.e., relevant onomasiological and cognitive phenomena (Grzeka 2002).

This feature represents a substantial difference from Dressler's (2005) approach which investigated MS transparency mostly in compounds, particularly in relation to his scale of MS transparency. By focusing on derivatives, the model of MS patterns enables the analysis of various semantic domains of the derived lexemes, emphasizing the radial architecture of the lexical system. An example of this is the morphosemantic field of *ruka* ('hand') or the root *ruk* as a lexical basis. As illustrated by some of the examples of the field listed below, the derived forms structuring the field have undergone diverse derivative and semantic processes, thus referring to a variety of entities or activities, all conceptually related to the base concept of 'hand'. Some of the lexemes structuring the morphosemantic field of the lexeme *ruka* 'hand' are ***ruč**nik* 'towel', ***ruč**ka* 'door handle', *iz**ruč**iti,* 'extradite', *u**ruč**iti* 'to deliver', *u**ruč**ak* 'handout', ***ruk**ovati* 'to manipulate', ***ruk**ovati se* 'to shake hand' and many others.[14] As the examples show, the derived forms can be semantically closely

[14] For examples of the morphosemantic field of the lexeme *ruka* in Croatian also see Katunar (2020).

or more remotely related to the lexical base. Derivatives like *uručiti* and *rukovati se* are conceptually closer to the lexeme *ruka* than the lexeme *izručiti* in which the conceptual relation between the noun *ruka* and the activity that is lexicalized is not salient anymore.[15]

Therefore, different types of morphosemantic relations and their radial structure (reflecting closer or more remote connections to the base word), as represented by the model of MS patterns, provide a more comprehensive understanding of paradigmatic lexical relations. This highlights the role of morphosemantic structures in the overall architecture of the lexicon. Previous approaches to MS transparency and paradigmatic relations in word-formation do not consider morphosemantic features as a principle of paradigmatic lexical organization that simultaneously integrates derivational and cognitive (primarily metaphor and conceptual contiguity) processes. As pointed out in the Introduction, the model of MS patterns offers a different perspective on motivation, which lies at the core of paradigmatic relations in the lexicon. Furthermore, from a psycholinguistic perspective, it provides new ways of exploring the interplay between the main paradigmatic relations – such as synonymy, antonymy, taxonomies – and morphosemantic structures in the speaker's mental organization of the lexicon. This means that the application of the model of MS patterns in the psycholinguistic research framework could provide insights into the way speakers shape their mental lexicon with respect to diverse paradigmatic lexical relations, which include morphosemantic features of the lexemes as well. From this point of view, it can be hypothesized that speakers would relate morphosemantically transparent lexemes (derivatives) to their lexical basis, i.e., the base word, with greater frequency than to their synonyms or co-hyponyms, as in the example of the adjective *sretan* 'happy', which more speakers (46.23%) closely related to the noun *sreća* 'happiness'[16] (the base word) than to its synonym *veseo* 'joyful' (30.85%) in our study.

MT and MS transparency are hard to disentangle in many cases. The application of the model of MS patterns in the psycholinguistic framework takes into account MT transparency features as well, thus pointing to their co-effects in building word semantics, as well as their interrelation and impact on the speakers' processing of words. In fact, our current study focused mostly on MT transparency, which was investigated to a lesser extent (see Section 2.1.) than MS transparency. However,

[15] Raffaelli (2016) introduces the notion of conceptual contiguity (as elaborated by Koch 1999) to explain conceptual relations between base words and their derivatives. For instance, the derivatives *uručiti* ('to deliver') and *rukovati se* ('to shake hands') are conceptually contiguous with the concept of *ruka* ('hand') lexicalized in the base word. In these cases, the concept of *ruka* remains preserved and salient. However, in the derivative *izručiti* ('to extradite'), the concept is no longer salient due to a metaphorical shift.

[16] This is the case of conceptual contiguity that allows for a closer relationship of the derivative (motivated word) to the base word.

the fact that all the stimuli in our study are derivatives presupposes morphosemantic relations as intrinsic to their semantic makeup. The model of MS patterns thus allows us to make predictions about the principles of lexical architecture and the way speakers shape their mental lexicon with respect to morphosemantic features of the lexemes, which is especially relevant for morphologically rich languages such as Croatian. Thus, the goal of our research was not the mere annotation of the MS transparency degrees in the various lexica but investigating what effects MS transparency has in assessing degrees of word relatedness and the paradigmatic structure of the lexicon.

In accordance with the questionnaire design for investigating MT transparency, only two types of MS transparency were taken into account in this study since, as already mentioned, the more fine-grained scale was impossible to incorporate into the current study design. The first category includes all examples in which the meaning of the stem and the meaning of the suffix are preserved, like in *kućica* 'house.DIM' < *kuća* 'house' + *-ica*. The second category consists of examples in which the meaning of the stem is remotely related to the meaning of the derivative, as in *maslina* 'olive' < *maslo* 'butter oil' + *-ina*. The lexemes within the first category are regarded as MS transparent derivatives, while the lexemes in the second category are regarded as MS non-transparent. This means that the notion of transparency is connected to the notion of compositionality and motivation, whereby there is a higher degree of compositionality in MS transparent than MS non-transparent derivatives. For example, while *kućica* denotes a small house and thus the concept 'house' remains salient in the derivative, *grašak* 'pea' does not denote a small bean, but a different type of legume.

The goal of introducing the investigation of MS transparency, along with MT transparency, was to gain a more comprehensive understanding of the interplay between morphological and semantic motivational pathways in the formation of the Croatian lexicon.

Based on the theoretical framework and previous research regarding MT and MS transparency in Croatian, it stands to reason that either MT transparency or MS transparency or both can have an impact on morphological awareness, that is, the ability of speakers to manipulate morphological structure of words (Apel, 2014). Therefore, the main aim of the study was to investigate whether MT and MS transparency could be considered as related to the accuracy of morpheme identification and word segmentation. Furthermore, a secondary aim was to investigate the place of morphological relations in the paradigmatic architecture of the Croatian lexicon. For this reason, different tasks for the study were developed, targeting our main research questions stated in the Introduction.

3. Methods

The study consisted of five tasks (see Procedure) designed to address our research questions, and each task targeted a different research question or one part of it:

RQ 1: Are base words and suffixes more accurately identified in MT transparent words? For this purpose, we designed Task 2 (identifying base words) and Task 4 (identifying suffixes).

RQ2: What role do morphological relations have in assessing word-relatedness?

a) Task 1 was designed as a ranking task to see how morphosemantic relatedness compares in rank to semantic relations of synonymy, antonymy, hyperonymy/hyponymy etc. We chose to begin the study with this task in order to avoid priming the speakers with other morphologically focused tasks later on and potentially influencing the ranking results.
b) Task 5 was designed to investigate whether MT and/or MS (non-)transparency can be tied to word-relatedness ratings on a scale from 1 to 5.

RQ3: Task 3 was designed to investigate whether speakers can identify those words that are derived and distinguish them from non-derived words. In a sense, its goal was to see if speakers can identify that some type of word-formation occurred and to examine if MT transparency contributes to this identification.

The tasks were originally developed for the purpose of this study. As we mentioned above (Section 2), we chose not to include full sentences with target words as stimuli to avoid syntactic context facilitating identification of MT and MS related words, therefore, each target word was shown in isolation.

3.1. Participants

Participants were all native speakers of Croatian with an age range of 20 – 65, mean age 31.98, 60.3% female, 39.7% male. A total of 38 participants listed secondary level education as the highest education level completed, half of which completed high school, and half a vocational school programme. The other 60 participants had a higher education degree, 31 from a non-philological and 29 from a philological study programme. A total of 101 participants completed the study but the results of 3 participants had to be excluded based on the low completion rate of the study. Results were analyzed for 98 participants, none of which reported having language difficulties or other difficulties (e.g., visual) influencing the completion of the tasks.

3.2. Materials

As the goal of the study was to investigate MT and MS transparency across lexical structures in Croatian, we chose a variety of derivationally frequent and productive suffixes, e.g., *-ica* (*učiteljica* 'teacher.F' < *učitelj* 'teacher.M'), *-telj* (*voditelj* 'leader' < *voditi* 'to lead'), *-ik* (*zlatnik* 'gold coin' < *zlatan* 'golden'), *-stvo* (*bogatstvo* 'wealth' < *bogat* 'wealthy'), *-ac* (*krastavac* 'cucumber' < *krastav* 'scabby'). These suffixes are productive in the sense that they are used in formation of many different derivatives and were extracted from earlier research on morphological segmentation of different parts of speech (see. Šojat et al. 2013 for verbs, Filko et al. 2017 for adjectives, Filko 2020 for nouns). Word stimuli (see Appendix) consisted of three major word classes in Croatian – nouns, verbs and adjectives – and the ratio of stimuli corresponded to the ratio of these word classes in the Croatian general lexicon as based on the total number of adjectives, nouns and verbs from the Croatian Language Portal dictionary[17], e.g., *kornjača* 'turtle.N', *raniti* 'to wound.V', *hladan* 'cold.ADJ'. In Task 1 and Task 3 morphologically unrelated words were added to address the specific questions of the tasks (preference for lexical semantic or morphological relatedness in Task 1 or disambiguation of morphologically complex vs. simplex words in Task 3). The order of presenting the stimuli was randomized within each task.

3.3. Procedure

As the study was conducted at the height of the Covid-19 pandemic, an online questionnaire was chosen as the optimal way to obtain data from various speakers of Croatian. Survey Monkey was used for designing and disseminating the study via an URL link, and participants who completed the study were asked to forward the link to other adult speakers of Croatian in the instructions. The data was collected over approximately two months in 2021.

The study consisted of five consecutive tasks in which participants were asked to identify, segment, or make judgments about derived words. Each task had its own set of instructions presented with each trial, and a practice example was listed within the instructions (see Appendix). Tasks requiring the production of language units (base words, suffixes) allowed participants to type their answers and participants could navigate to the next question at their own pace. The average time needed to complete the study was around 30 minutes.

In Task 1 participants were asked to rank a set of five words based on them being most related (rank 1) or least related (rank 5) to the stimulus word. The goal of the task was to explore the overall place of morphological relatedness within the set of lexical relations for Croatian, and to obtain both quantitative and qualitative insights into

[17] Available at hjp.znanje.hr.

preferences for semantic relations. Each word in the list was related to the stimulus word based on a lexical semantic relation or a morphological relation, e.g., *bježati* 'escape, run away' (stimulus) – *bijeg* 'escape.N' (morphological relation) – *izmicati* 'to elude' (synonym) – *loviti* 'to chase; to hunt' (antonym) – *kretati se* 'to move' (hyperonym) – *trčati* 'to run' (semantic field). There were 15 trials, each containing five morphologically or semantically related words presented in a randomized order. Task 1 was examined with respect to rank percentiles assigned to a particular relation.

Task 2 investigated the ability of speakers to identify and produce base words related to stimuli words, e.g., *junački* 'heroic' < *junak* 'hero', *maslina* 'olive' < *maslo* 'butter oil', *čajnik* 'teapot' < *čaj* 'tea'. There were 30 stimuli items presented consecutively, half of the stimuli were MT transparent words, e.g., *čajnik* 'teapot' < *čaj* 'tea', while half were MT non-transparent words, e.g., *junački* 'heroic' < *junak* 'hero'. Responses were coded for accuracy, e.g., *grašak* 'peas' < *grah* 'beans' (accurate); *biži* 'peas. dialectal', *mahuna* 'beanstalk' (inaccurate). In Task 2 (unlike Task 3 and 4), more than one base word was coded as accurate to allow for the possibility of multiple segmentations (when these were listed by participants as answers), e.g., *krastavac* 'cucumber' < *krasta* 'scab' or *krastav* 'scabby'. Overall, 10 stimuli items had multiple answers coded as correct, most with only two answers (e.g., *vozilo* 'vehicle' < *voz* '(animal) wagon; archaic' and *voziti* 'to drive' (accurate); *vožnja* 'a ride', *prometalo* 'vehicle, synonym' (inaccurate); and *svitak* 'scroll' was accepted with 4 possible segmentations due to alternating base word spellings, *sviti/saviti* 'to fold. PF' and *svinuti/savinuti* 'to bend.PF'.

Task 3 was designed to investigate whether speakers of Croatian can distinguish between derived and non-derived (simplex) words, and was presented as a checklist identification task, in which participants were asked to identify (mark) those words that are formed from other words, i.e., derivatives. There were 26 stimuli items presented as a list in a randomized order; ten were MT transparent words, e.g., *vidjeti* 'to see' < *vid* 'sight.N', eight MT non-transparent words, e.g., *lišće* 'leaves' < *list* 'leaf' and eight simplex words, e.g., *žlica* 'spoon', *nebo* 'sky'. The task was coded for accuracy and error rates were examined (more in Results).

Task 4 was a mirror-image version of Task 2, as it asked participants to identify and segment suffixes used to form a stimuli derivative word, e.g., *raniti* 'to wound' < *rana* 'wound.N' + ***-iti***, *kraljica* 'queen' < *kralj* 'king'+ ***-ica***, *lišće* 'leaves' < *list* 'leaf' + ***-je***, *viteštvo* 'knighthood' < *vitez* 'knight'+ ***-stvo***. The participants were required to write only the suffixes, and not the stems or base words. There were 22 stimuli items presented, 12 MT transparent and 10 MT non-transparent words. Unlike Task 2, where multiple accurate answers were sometimes recorded and coded as such, each stimulus in Task 4 was coded as having one possible accurate answer – the form of the suffix before any morphonological changes occurred. This way the results could show if MT transparency had any effect on suffix identification.

While Tasks 2, 3 and 4 focused on MT transparency, Task 5 was added to get a preliminary insight into the relationship between MT and MS transparency. In Task 5 participants were required to evaluate word relatedness between pairs of words on a scale from 1 to 5. Sixteen pairs of words were divided into four categories: 1) MT and MS transparent, e.g., *kuća* 'house' > *kućica* 'house.DIM' 2) MT transparent – MS non-transparent, e.g., *prijati* 'be pleasant' > *prijatelj* 'friend' 3) MT non-transparent – MS transparent, e.g., *pjevati* 'to sing' > *pjesma* 'song', 4) MT and MS non-transparent, e.g., *kornjača* 'turtle' < *kora* 'bark'.

4. Results

4.1. MT transparency and the identification and segmentation of derivatives

Answers in Tasks 2, 3 and 4 were coded for accuracy, and a chi-square test 2x2 table (with Yates correction) was used. The results show a correlation between MT transparency and accuracy in identifying suffixes (Task 4), as well as distinguishing between derived and simplex words (Task 3). In Task 4 suffixes were more accurately identified in MT transparent than MT non-transparent words ($\chi^2 = 234.48$, $df = 1$, $p < .01$), and in Task 3 MT transparent words were more accurately identified as motivated (derived) than MT non-transparent words ($\chi^2 = 166.35$, $df = 1$, $p < .01$). In Task 3 the overall error rate for identifying unmotivated (simplex) words as motivated was very low (2.55%), which is also evident from the error rates for individual unmotivated words in Figure 1.

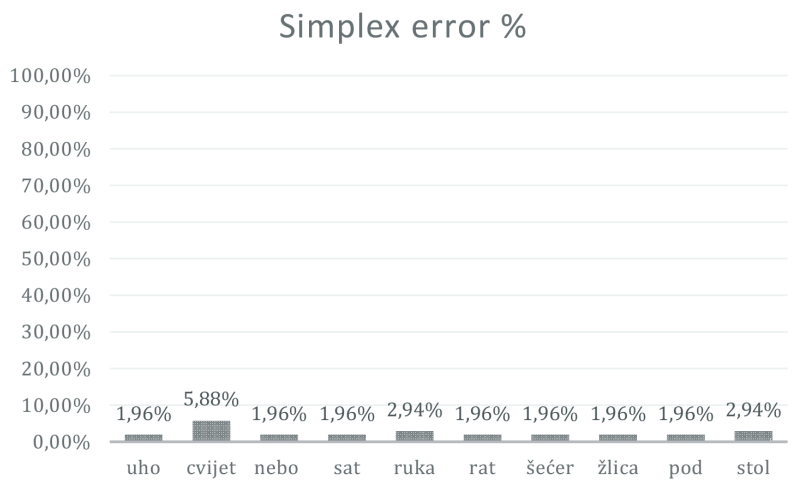

Figure 1. Task 3 – Error rates for individual unmotivated words. As the percentiles show, speakers have a very low error rate when it comes to simplex words inaccurately being identified as motivated.

However, in Task 2 MT transparency does not seem to be significantly related to accurate base word production ($\chi^2 = 2.30$, $df = 1$, $p > .05$). To illustrate the point, the most accurately identified base words (over 80 participants listed them as answers) were six MT non-transparent words, *junak* 'hero' > *junački* 'heroic, non-MT', *mozak* 'brain' > *mozgati* 'to brainstorm, non-MT', *grah* 'beans' > *grašak* 'peas, non-MT' and *živac* 'nerve' > *živčan* 'nervous, non-MT', and three MT transparent words, *kuća* 'house' > *kućica* 'house.DIM, MT', *čaj* 'tea' > *čajnik* 'teapot, MT', *škola* 'school.N' > *školski* 'school.ADJ, MT'.

4.2. Assessing word relatedness

Tasks 1 and 5 were not coded for accuracy but were used instead to gain an insight into tendencies for morphological or semantic preferences regarding ranking or evaluating morphological relatedness.

In Task 1 results were analyzed as percentiles of each rank with respect to each relation to the target word. For example, target word *ljutnja* 'anger' had 60% of participants assign rank 1 (most related) to its morphologically related base word *ljut* 'angry';[18] compared to 20% for synonym *srdžba* 'fury', 3% for antonym *smirenost* 'calmness', 9% for hypernym *osjećaj* 'feeling' and 10% for semantic field member (i.e., word belonging to the same semantic domain) *bijesniti* 'to rage'. Rank 5 (least related) for the same word was assigned as follows: 3% for the base word, 5% for the synonym, 85% for the antonym, 5% for the hypernym and 2% for semantic field member. Different words in the task varied according to rank assignment, thus the results of this task were examined for preference of types of paradigmatic relations assigned rank 1. For rank 1, participants prefer synonymy and morphological motivation when assessing relatedness between words (more so than other semantic criteria, such as antonymy or hyperonymy/hyponymy), as 8 out of 15 words were assigned rank 1 based on synonymy and 6 based on morphological relatedness.

Task 5 required participants to rate the relatedness of words on a scale from 1 (not related) to 5 (strongly related). The average rating on the scale was used to assess whether there is any tendency to rate either MS or MT transparent words as more related with respect to MS or MT non-transparent words. In Task 5 the results show that MS transparency has an effect on assessing word relatedness, as in Figure 2, but MT transparency seems to show no effect.

[18] This is also an example of conceptual contiguity, which allows for a strong morphosemantic relationship, making morphosemantic paradigmatic relations generally more salient than other paradigmatic relations

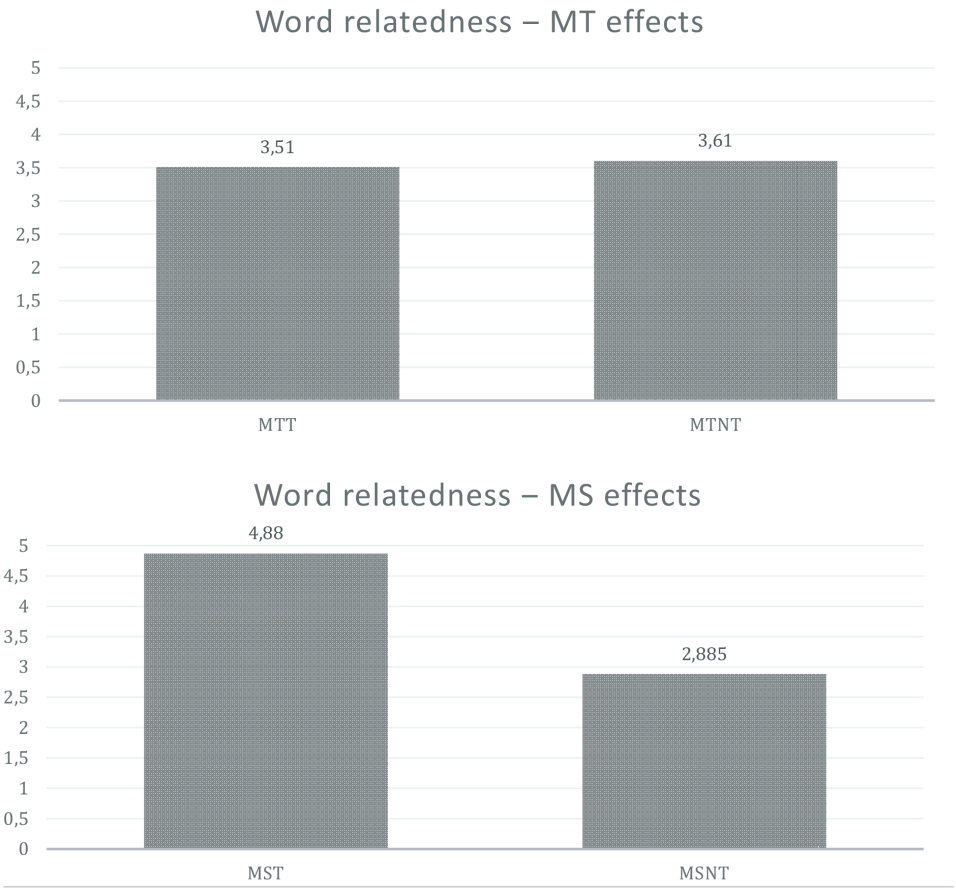

Figure 2. Mean average of word relatedness ratings for the categories of morphotactically transparent (MTT) – morphotactically non-transparent (MNT) and morphosemantically transparent (MST) – morphosemantically non-transparent (MSNT) words. The same dataset was coded for the two variables separately.

4.3. Education and word segmentation accuracy

Although MT transparency was taken as a starting point in designing Tasks 2 – 4, and we did not control for an equal split between participants based on level of education as this was not the focus of the study, one relevant result when it comes to participant data seems to be the type of education completed. Participants with a philology degree are more accurate in identifying suffixes (Task 4: $\chi^2 = 4.56$, $df = 1$, $p < .05$). Also, participants with a philology degree are more accurate in identifying MT transparent motivated words (Task 3) than participants with a non-philological degree ($\chi^2 = 10.68$, $df = 1$, $p < .01$), but there are no significant differences in accu-

racy when both groups try to identify MT non-transparent words as motivated (χ^2 = 0.0006, *df* = 1, *p* > .05). Identification of base words (Task 2) shows no significant effect of education between the two groups (χ^2 = 0.51, *df* = 1, p > .05). When it comes to secondary education level groups, participants with a high school education are more accurate than participants with completed vocational school programmes:

1) in identifying suffixes (Task 4: χ^2 = 14.06, *p* < .01),

2) in distinguishing MT transparent and MT non-transparent from unmotivated words (Task 3: χ^2 = 31.97, *df* = 1, *p* < .01; χ^2 = 6.70, *df* = 1, *p* < .01),

3) in producing base words of derivatives (Task 2: χ^2 = 11.78, *df* = 1, *p* < .01).

However, when it comes to the effect of the level of education on the accuracy of identifying and segmenting derived words, results show no overall difference in Tasks 2 – 4 with regards to secondary or higher education (χ^2 = 0.16, *df* = 1, *p* > .05; χ^2 = 0.39, *df* = 1, *p* > .05; χ^2 = 0.23, *df* = 1, *p* > .05; χ^2 = 0.25; *df* = 1; *p* > .05).

5. Discussion

As this study was the first of its kind for Croatian, it provided some important insights into the role of morphology in structuring Croatian lexical relations.

First and foremost, the results clearly indicate that MT transparency facilitates identification of suffixes as separate morphological units, as investigated by Task 4 where participants were asked to identify and write down suffixes of target words. One possible explanation is a clear delineation of the stem and the suffix in the morphological structure of MT transparent words, i.e., *mere juxtaposition* in Dressler's terminology. For example, noun *kralj* 'king' is recognizable as a separate lexical unit and it is clearly separable from the morphological structure of the noun *kraljica* 'queen', thus leaving *-ica* as the unambiguous candidate for a suffix. Conversely, in examples in which there is a sound change at the morpheme boundary (i.e., in MT non-transparent words), and particularly in examples where fusion occurs at the stem – suffix boundary (e.g., *junak* 'hero'+ *-ski* > *junački* 'heroic', *posuda* 'bowl' + *-je* > *posuđe* 'dishes; pots and pans'), speakers cannot distinguish between the phonological composition of the stem and the suffix, i.e., a clear suffix candidate does not present itself easily.

A similar interpretation could be offered for results in Task 3, distinguishing between derived and non-derived words, where MT transparent words contain more easily identifiable base words and suffixes (see example *kralj – kraljica* above) and thus present themselves as more unambiguous candidates for derivatives. This means that speakers can identify base words and suffixes in MT transparent words, which serve as cues for identifying that derivation has occurred.

Though this was not the primary focus of the study, an analysis of the types of errors speakers make when segmenting suffixes in Task 4 seems to support this interpretation. For example, in the lexeme *bolnički* 'hospital.A' (< *bolnica* 'hospital' + *-ski*), the ending of the stem (*bolni**c***-) fuses with the beginning of the suffix (*-**s**ki*) resulting in the fused form *bolni-**č**-ki*. From a morphological point of view, *-č-* should be considered as part of the stem (cf. Filko 2020, 74–76), however, 75.49% of the speakers listed *-čki* as the suffix. Likely reasons for that are a) forms with analogous sound changes, e.g., *junak* 'hero' > *junački* 'heroic' and b) the existence of the suffix *-ački*, with a more or less analogous form, e.g., *Zagreb – zagrebački* 'Zagrebian.' Only around 3% of the speakers listed the target answer, i.e., the suffix *-ski*, or listed its allomorph *-ki*, thus correctly assigning the fused phoneme *-č-* to the stem. Finally, around 10% of the speakers left this question unanswered (as they were instructed to in the description of the task), showing that they cannot recognize the suffix in the MT non-transparent word.

Conversely, conceptual and formal autonomy of the lexical base (or base word) may explain the results from Task 2, in which MT transparency did not show a significant relation to accuracy in identifying lexical bases/base words. However, results indicate that there are other factors at play, most notably semantics. Although it would be expected that MT non-transparent lexemes would be more difficult to process morphosemantically, i.e., to semantically relate them to their lexical base, some of our results show the opposite. The examples of *pleter* 'interlace; braided ornament' and *grašak* 'pea' show that speakers have consistently related the nouns to the base words, i.e., the verb *plesti* 'to knit' (59%) and the noun *grah* 'beans' (81%). These results show semantics' impact on the paradigmatic relations in word-formation, i.e., the lexical organization of morphosemantically related words. This should not be overlooked even when the relations between the stem and the derivative can be blurred on the morphotactic level. This means that conceptual contiguity has a significant impact on the semantic features of morphosemantic lexical structures, thereby strengthening the relationship between base words and derivatives, despite MT non-transparency, e.g., *mozak* 'brain' > *mozgati* 'to think over; to brainstorm', as predicted by the model of MS patterns (see Section 2). Importantly, it must be said that we accepted more than one answer as correct when there was a possibility of multiple segmentations, e.g., *krastavac* 'cucumber' < *krasta* 'scab' or *krastav* 'scabby', which is also in accordance with the model of MS patterns.

Tasks 1 and 5 can be considered secondary with respect to the main conclusions of the study as there are limitations to their design. Though ranking results from Task 1 undoubtedly support the hypothesis that morphosemantic paradigmatic relations have an important role in structuring Croatian lexical relations (especially when conceptual contiguity is operative in the relationship between a base word and its derivative), as they were one of the primary choices for speakers besides

synonymy, a forced ranking paradigm should be supplemented with other types of tasks in the future (for example, rating scales). As we mentioned in Section 2, morphosemantic features play an important role in paradigmatic lexical relations in Croatian as a morphologically rich language and they should be further investigated experimentally. Similarly, future research could focus on disentangling the notions of MT and MS transparency as interrelated, but in some cases still distinct notions. As exemplified by Task 5 (rating word relatedness on a scale from 1 to 5), it seems that MS transparency is the driving factor behind word-relatedness assessments, which is also supported by the MS patterns model. However, MT and MS transparency are hard to disentangle in many cases, thus limiting the stimuli selection needed for an extensive experimentally and statistically comprehensive design.

Within the current study design, a question arises as to the nature of the identification ability of speakers. Since the study was an online questionnaire, which required speakers to make conscious metalinguistic judgments at their own pace, the results could partly be a product of the study design itself and reflect (primarily) morphological awareness, while other aspects of morphological processing should be investigated with additional experimental designs (measuring reaction time, priming effects and so forth). The relation of type of education (philology/non-philology; high school/vocational school) to suffix and derived words identification accuracy in Tasks 3 and 4 indicates that a more language-focused type of education could have an effect on morphological awareness. However, this should be investigated more thoroughly in future studies targeting this question specifically from the onset. Nevertheless, to the best of our knowledge, this is the first study showing that the type of education is a possible predictor of morpheme segmentation accuracy for Croatian speakers. Of course, based on the results reported for English, such as Carlisle (2000), future studies on Croatian should investigate morphological awareness in the earlier stages of language development as well. The results of the present study show that it remains constant in secondary and higher education levels, as there were no significant differences observed between these two levels in any of the tasks.

Finally, unlike previous MS and MT transparency research, which takes psycholinguistic models of morphological processing and reading/ reading comprehension as the starting point, this study was based on theoretical (structuralist and cognitive linguistic) approaches to paradigmatic lexical relations, specifically the model of MS patterns. As suggested by this model, MT and MS transparency are relevant factors in organizing paradigmatic relations in the Croatian lexicon, and the results presented in this study add to this view of the interrelatedness of morphological and semantic structures.

6. Conclusion

The research presented in this paper was the first step towards a comprehensive investigation of MT and MS transparency in Croatian. Apart from the complete lack of experimental research on MT and MS transparency in Croatian, theoretical accounts of these phenomena are also scarce. Thus, for the initial questionnaire, we disregarded fine-grained scales present in the literature and decided to divide lexemes into two broad groups: transparent and non-transparent. Regarding MT transparency, transparent lexemes are lexemes without any morphonological changes at morpheme boundaries, while MT non-transparent lexemes have one or more morphological changes in their morphological structure. MS transparent lexemes are those with a compositional meaning, i.e., lexemes where the meaning of both the base word and the affix is preserved; unlike the MS non-transparent lexemes, where either the base word or the affix or the derived lexeme have undergone a shift in meaning.

The main aim of the study was to investigate whether morphology and morphosemantics play a significant role in processing Croatian derivatives, i.e., motivated words, as opposed to unmotivated words. The results of the study conducted on 98 native speakers of Croatian show that:

a) MT transparency is associated with the recognition of the suffix as a non-autonomous unit while the base word is equally recognizable in both MT transparent and MT non-transparent lexemes. We can postulate at least two possible reasons for this: 1) base words are autonomous units and thus more salient in the mental lexicon; 2) base words are phonologically longer than affixes and thus easier to recognize. A supplementary study would be needed to investigate whether speakers are successful in stem recognition, which is phonologically altered in MT non-transparent words. However, this was not part of the current study design. MT transparency also affects the recognition of motivated lexemes, i.e., speakers more accurately detect MT transparent derivatives as being derived, while they more often consider MT non-transparent derivatives as non-derived

b) MS transparency affects the ratings of morphological relatedness, i.e., base word – derivative pairs with MS transparent derivatives preserving conceptual salience (e.g. *kuća* 'house' – *kućica* 'house, diminutive') are considered more closely related than the pairs with MS non-transparent derivatives (e.g. *prijati* 'to please' – *prijatelj* 'friend')

c) morphosemantic structures play an important role in the overall paradigmatic organization of the Croatian lexicon, affecting both the ranking of word relations (compared to other paradigmatic lexical relations, such as synonymy, semantic fields etc.) and the assessment of word relatedness. This indicates that they should be investigated further as part of the group traditionally considered paradigmatic lexical relations

d) there is no overall difference in the accuracy of identification of morphological units between speakers with secondary and higher education. However, type of education has an effect in some cases. Speakers with a high school education are more accurate in the identification of morphological units than speakers with vocational schools, and speakers with a philological background are more accurate in recognizing MT transparent motivated lexemes in Task 3 (see Appendix) than speakers with a non-philological background.

These results show that morphosemantic structures are important to the structuring of lexical relations in Croatian and that MT and MS transparency are important for morphological processing of Croatian derivatives. However, a more systematic study of morphological relations as competing with lexical semantic relations (e.g., synonymy) is needed in future research. This opens new avenues for morphological studies of Croatian and future psycholinguistic approaches to Croatian morphology.

Finally, a question arises as to the nature of the identification ability – is it explicit metalinguistic knowledge (an effect of education) or implicit speaker knowledge? Research into earlier stages of language acquisition is needed to answer that question, although the results of this study lead us to conclude that morphological awareness is consistent from secondary education onwards.

References

Apel, Kenn. 2014. A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, 34(3). 197–209.

Babić, Stjepan. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. 3., poboljšano izdanje. Zagreb: Hrvatska akademija znanosti i umjetnosti: Nakladni zavod Globus.

Bagasheva, Alexandra. 2020. Paradigmaticity in compounding. In Fernández-Domínguez, Jesús, Bagasheva, Alexandra, Clares, Cristina Lara (eds.), *Paradigmatic relations in word formation*, 21–48. Leiden/Boston: Brill.

Baldinger, Kurt. 1984. *Vers une sémantique moderne*. Paris: Klincksieck.

Carlisle, Joanne F. 2000. Awareness of the structure and meaning of the morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal* 12. 169–190.

Diependaele, Kevin, Sandra Dominiek and Jonathan Grainger. 2009. Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition* 37(6). 895–908. 10.3758/MC.37.6.895

Dressler, Wolfgang U. 1985. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21(2). 321–337.

Dressler, Wolfgang U. 2005. Word-formation in Natural Morphology. In Pavol Štekauer, Lieber, Rochelle (eds.), *Handbook of Word-Formation, Studies in Natural Language and Linguistics Theory*, 267–284. Berlin: Springer.

Feldman, Laurie Beth, Dragana Barac-Cikoja and Aleksandar Kostić. 2002. Semantic aspects of morphological processing: Transparency effects in Serbian. *Memory & Cognition* 30(4). 629–636. 10.3758/BF03194964

Feldman, Laurie Beth and Jennifer A. Stolz. 1995. The Role of Orthographic and Semantic Transparency of the Base Morpheme in Morphological Processing. In Feldman, Laurie Beth (ed.) *Morphological aspects of language processing*, 109–129. Hillsdale, NJ: Erlbaum.

Filko, Matea. 2017. Expansion of the derivational database for Croatian. In Litta, Eleonora, Passarotti, Marco (eds.), *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, 27–37. Milano: EDUCatt.

Filko, Matea, Krešimir Šojat and Vanja Štefanec. 2020. The Design of Croderiv 2.0. *The Prague Bulletin of Mathematical Linguistics* 115. 83–104. 10.14712/00326585.006

Filko, Matea. 2020. *Unutarleksičke i međuleksičke strukture imeničkoga dijela hrvatskoga leksika*. Neobjavljena doktorska disertacija. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.

Fradin, Bernard. 2020. Characterazing Derivational Paradigms. In Fernández-Domínguez, Jesús, Bagasheva, Alexandra, Clares, Cristina Lara (eds.), *Paradigmatic relations in word formation*, 49–84. Leiden/Boston: Brill.

Frisson, Steven, Niswander-Klement, Elizabeth, Pollatsek, Alexander. 2008. The role of semantic transparency in processing of English compounds. *British Journal of Psychology* 99(1). 87–107. doi: 10.1348/000712607X181304

Gagné, Christina, Marchak, Kristian, Spalding, Thomas. 2010. Meaning predictability and compound interpretation: A psycholinguistic investigation. *Word structure* 3(2). 234–251.

Geeraerts, Dirk. 1997. *Diachronic Prototype Semantics – A Contribution to Historical Lexicology*. Oxford: Clarendon Press.

Guiraud, Pierre. 1967. *Structures étymologiques du lexique français.* Paris: Larousse.

Grzega, Joachim. 2002. Some Thoughts on a Cognitive Onomasiological Approach to Word-Formation with Special Reference to English. *Onomasiology Online* 3. 1–29.

Hay, Jennifer. 2002. From Speech Perception to Morphology: Affix Ordering Revisited. *Language* 78(3). 527–55.

Hay, Jennifer, Baayen, Harald. 2002. Parsing and productivity. In Booij, Geert, Van Marle, Jaap (eds), *Yearbook of Morphology 2001. Yearbook of Morphology*, 203–235. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-3726-5_8

Hay, Jennifer, Plag, Ingo. 2004. What Constrains Possible Suffix Combinations? On the Interaction of Grammatical and Processing Restrictions in Derivational Morphology. *Natural Language & Linguistic Theory* 22(3). 565–96.

Katunar, Daniela. 2020. Cognitive Semantics. In Greenberg, Marc L. (ed.), *Encyclopedia of Slavic Languages and Linguistics Online*. Brill. https://doi.org/10.1163/2589-6229_ESLO_COM_032520

Koch, Peter. 1999. *Frame and Contiguity: On the cognitive bases of metonymy and certain types of word formation.* Panther, Klaus-Uwe, Radden, Günter (eds), *Metonymy in Language and* Thought, 139–167. Amsterdam, Philadelphia: John Benjamins.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Volume 1. Theoretical Prerequisites*. Standford: Standford University Press.

Langacker, Ronald W. (2000). *Grammar and Conceptualization*. Berlin and New York: Mouton de Gruyter.

Libben, Gary. 1998. Semantic transparency in the processing of compounds: Consequence for representation, processing, and impairment. *Brain and Language*, 61(1). 30–44. https://doi.org/10.1006/brln.1997.1876

Libben, Gary and Silke Weber. 2014. Semantic transparency, compounding, and the nature of independent variables. In Rainer, Franz, Dressler, Wolfgang U., Gardani, Francesco, Luschützky, Hans Christian (eds.) *Morphology and meaning*, 205–222. Amsterdam: Benjamins.

Mahony, Diana, Maria Singson and Virginia Mann. 2000. Reading ability and sensitivity to morphological relations. *Reading and Writing: An interdisciplinary Journal* 12. 191–218.

Manova, Stela (ed.). 2015. *Affix ordering across languages and frameworks*. New York, NY: Oxford University Press.

Marković, Ivan. 2012. *Uvod u jezičnu morfologiju*. Zagreb: Disput.

Marković, Ivan. 2013. O najvećim (i) mogućim hrvatskim riječima. In Blažetin, Stjepan (ed.), *XI. međunarodni kroatistički znanstveni skup,* 43–58. Pečuh: Znanstveni zavod Hrvata u Mađarskoj.

Mattiello, Elisa, Dressler, Wolfgang U. 2018. The Morphological Transparency/Opacity of Novel English Analogical Compounds and Compound Families. *Studia Anglica Posnaniensia* 53(2018). 67–114.

Plag, Ingo. 2002. The role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. In Booij, Geert, Van Marle, Jaap (eds), *Yearbook of Morphology 2001. Yearbook of Morphology*, 285–314. Springer, Dordrecht.

Raffaelli, Ida. 2012. The Conceptual Category of Light in Croatian: a diachronic perspective. In Brdar, Mario, Raffaelli, Ida, Žic Fuchs, Milena (eds), *Cognitive Linguistics between Universality and Variation*, 383–410. Cambridge Scholar Publishing, Cambridge.

Raffaelli, Ida. 2013. The model of morphosemantic patterns in the description of lexical architecture. *Lingue e linguaggio* 1. 47–72.

Raffaelli, Ida. 2015. *O značenju – Uvod u semantiku.* Zagreb: Matica hrvatska.

Raffaelli, Ida. 2016. Morfosemantička obilježja percepcijskih glagola u hrvatskom. In Botica, Stipe, Nikolić, Davor, Tomašić, Josipa, Vidović Bolt, Ivana (eds), *Zbornik radova Šestog hrvatskog slavističkog kongresa,* 375–387. Hrvatsko filološko društvo, Zagreb.

Raffaelli, Ida, Kerovec, Barbara. 2008. Morphosemantic fields in the analysis of Croatian vocabulary. *Jezikoslovlje* 9(1–2). 141–169.

Schwaiger, Sonja et al. 2017. Scaling morphosemantic transparency/opacity: A corpus-linguistic and acquisitionist study of German diminutives. *Yearbook of the Poznań Linguistic Meeting* 3. 141–153. 10.1515/yplm-2017-0007

Šojat, Krešimir, Matea Srebačić and Marko Tadić. 2012. Derivational and Semantic Relations of Croatian Verbs. *Journal of Language Modelling* 00(1). 111–142.

Šojat, Krešimir, Matea Srebačić and Vanja Štefanec. 2013. CroDeriV i morfološka raščlamba hrvatskoga glagola. *Suvremena lingvistika* 75(1). 75–96.

Štekauer, Pavol. 2005. Onomasiological Approach to Word-Formation. In Štekauer, Pavol, Lieber, Rochelle. (eds.), *Handbook of Word-Formation*. 207–232. Springer, Dordrecht.

Štekauer, Pavol 2014. Derivational paradigms. In Lieber, Rochelle, Štekauer, Pavol. (eds.), *The Oxford Handbook of Derivational Morphology*. 354–369. Oxford University Press, Oxford.

Talamo, Luigi, Chiara Celata and Pier Marco Bertinetto. 2016. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure* 9(1). 72–102.

Appendix 1: The questionnaire setup

Task 1. Poredajte sljedeće riječi prema tome koju najviše povezujete sa zadanom riječi (1) do one koju najslabije povezujete sa zadanom riječi (5). [Order the following words according to most (1) and least (5) related to the given word.]

prsten 'finger'

prst 'finger', burma 'wedding ring', zaprositi 'propose', narukvica 'bracelet', ruka 'hand'

umoran 'tired'

umoriti se 'to tire', iscrpljen 'be fatigued', odmoran 'be rested', krevet 'bed', napor 'exertion'

raniti 'to wound'

rana 'a wound', povrijediti 'to hurt', iscijeliti 'to heal', ubosti 'to stab', ozljeda 'injury'

…

Task 2. Napišite riječ od koje je nastala navedena riječ (npr. riječ *prijateljica* nastala je od riječi *prijatelj*). Ako ne prepoznajete riječ od koje navedena riječ dolazi, ostavite prazno. [Write down the word used to form the word listed below (e.g., the word *prijateljica* 'friend.F' is formed from the word *prijatelj* 'friend.M'). If you do not recognize the word used to form the word listed below, leave blank.]

mozgati 'to brainstorm'

junački 'heroic'

ljepota 'beauty'

…

Task 3. Među navedenim riječima prepoznajte i označite one riječi koje su tvorene od drugih riječi, npr. riječ *školski* tvorena je od riječi *škola* nastavkom *ski*. Riječi koje nisu nastale od drugih riječi ostavite neoznačenima (npr. riječ *led* nije tvorena od neke druge riječi). [Among the words listed please check those which come from other words, e.g., the word *školski* 'school.ADJ' is formed from *škola* 'school.N' by the affix *-ski.* Leave unchecked words that are not formed from other words (e.g., the word *led* 'ice' was not formed from another word).]

uho 'ear'

bježati 'run away'

tenisač 'tennis player'

…

Task 4. Napišite nastavak kojim je tvorena sljedeća riječ (npr. riječ *školski* tvorena je od riječi *škola* nastavkom *ski*). [Write the suffix used to form the following word (e.g., the word *školski* 'school.ADJ' is formed from *škola* 'school.N' with the affix *-ski*.)]

brdovit 'hilly, mountainous'

viteštvo 'knighthood'

raniti 'to wound'

…

Task 5. Označite stupanj povezanosti koji osjećate među navedenim parovima riječi. [Mark the degree of relatedness that you perceive between the listed word pairs.]

grah – grašak 'beans – peas'

prijati – prijatelj 'to please – friend'

bijeg – bježati 'an escape – to escape'

…

JEZIKOSLOVLJE
25.2 (2024): 269-296

ISTRAŽIVANJA MORFOTAKTIČKE I MORFOSEMANTIČKE TRANSPARENTNOSTI – PRIMJER SUFIKSACIJE U HRVATSKOM

U radu se predstavlja prvo istraživanje morfotaktičke (MT) i morfosemantičke (MS) transparentnosti u hrvatskome jeziku. Istraživanje je provedeno putem *online* upitnika na platformi Survey Monkey i sastojalo se od niza zadataka u kojima su sudionici trebali identificirati morfeme ili segmentirati riječi s različitim razinama MT i MS transparentnosti te dati prosudbe o povezanosti riječi na temelju morfoloških ili semantičkih kriterija. Sve motivirane riječi iz istraživanja nastale su sufiksacijom i pripadaju općem leksiku hrvatskoga jezika. U istraživanju je sudjelovao 101 odrasli izvorni govornik hrvatskoga jezika. Studija je osmišljena kako bi odgovorila na nekoliko istraživačkih pitanja: 1) hoće li govornici lakše izdvajati sufikse i polazišne riječi u MT transparentnim riječima; 2) koja je uloga morfoloških odnosa u procjeni paradigmatske povezanosti između riječi; 3) mogu li govornici razlikovati tvorenice od netvorbenih riječi. Rezultati potvrđuju da su i MT i MS transparentnost važne za morfološku obradu hrvatskih izvedenica.

Ključne riječi: morfotaktička transparentnost, morfosemantička transparentnost, model morfosemantičkih obrazaca, sufiksacija, hrvatski

Authors' addresses:

Ida Raffaelli
Faculty of Humanities and Social Sciences
HR – 10 000 Zagreb, Ivana Lučića 3
ida.raffaelli@ffzg.unizg.hr

Matea Filko
Faculty of Humanities and Social Sciences
HR – 10 000 Zagreb, Ivana Lučića 3
matea.filko@ffzg.unizg.hr

Daniela Katunar
Faculty of Humanities and Social Sciences
HR – 10 000 Zagreb, Ivana Lučića 3
dkatunar@ffzg.unizg.hr