

Word-formation type, its reinterpretations, and possible equivalents¹

Michał Rzepiela, Institute of Polish Language of The Polish Academy of Sciences,
Kraków

This paper discusses terminology used to classify the groups of derivatives characterized by some common formal and semantic features and especially those coined with one specific affix. Special attention is given to the concept of word-formation by Czech scholar Miloš Dokulil and the phrase word-formation type introduced by him. One points out a strict hierarchical order of the terms relating to the products of word-formation in Miloš Dokulil's framework and demonstrates how the phrase word-formation type was reinterpreted by other scholars regarding the exploitation of electronic corpora by František Štichá and to the onomasiological theory by Pavol Štekauer. The terms microstructure (lexicale) and semantische Nische, employed in French and German studies of word-formation, respectively, are comparatively recalled. Finally, attention is focused on the phrase lexico-semantic class and its use as usually encountered in computational linguistics.

Keywords: word-formation terminology, classification of lexical items, word-formation type, lexico-semantic class, computational linguistics, Latin

1. Introduction

The phrase *word-formation type* (Czech *slovo tvorný typ*; hereafter WFT), commonly used in the studies of word-formation (WF) in Central European countries, especially in the Czech Republic, Slovakia and Poland, seems to have a rather local extension and might not be understandable to scholars formed by currents outside of the Prague school of structural linguistics. It was coined as part of an original terminological apparatus by Dokulil (1962) to describe WF rules in modern Czech. However, for me, the inspiration to examine Dokulil's terminology came not from Czech as much as from Latin. Some years ago, when I was doing my PhD thesis research on suffixation in medieval Latin (published as an article in 2005), I paid more attention to Polish studies on WF, and I then realized how strongly the approach of Polish linguists, both in theoretical and terminological aspects, was influenced by the ideas of Dokulil. In this way, I hit for the first time Dokulil's pattern of derivatives classification, which I then hoped to apply to Latin.

For me, particularly interesting in his concept – let us remember, based on onomasiological presumptions and structuring lexical items according to the degree of meaning's abstractness and generality² – is the level at which formal (morphological) and semantic criteria meet. Dokulil (1962), as the most general term combining these criteria, proposes *word-formation category*. Such a category within nouns, for example, is constituted by names of professions. Lexemes such as *decretista* 'expert in canon law' (← *decretum*), *forestarius* 'forest ranger' (← *foresta*), *mensator* 'carpenter' (← *mensa*) may be enclosed into this category in medieval Latin. They represent the identity of semantics and WF bases (all are desubstantive) but are coined with different suffixes. However, when derivatives

¹ I wish to thank the anonymous reviewer for proofreading my paper and for his/her precious remarks, as well as bibliographical suggestions, which have significantly widened my insight into the discussed issue.

² We leave aside presentation of Dokulil's onomasiological theory since it has been recently broadly discussed, e.g. Štekauer (2005), Panocová (2015).

characterized by semantics and WF base are, in addition, coined with the same suffix, e.g. Lat. *-tor*; they are formed according to terminology used by Dokulil: WFT, as illustrated in the Latin deverbal names of professions: *braxator* ‘brewer’ (← *braxare* ‘to brew’), *falcastrator* ‘mower’ (← *falcastrare* ‘to mow’), *impressator* ‘printer’ (← *impressare* ‘to print’). Moreover, to mark any slight semantic differentiation between the lexemes belonging to the same type, it is possible to divide into further subtypes. For example, Latin nouns using *-ista* to designate musicians of very strict specializations (*clavichordista* ‘clavichordist’, *lutnista* ‘lutenist’, *organista* ‘organist’) can be interpreted as a *subtype* of the WFT of desubstantive names of professions using *-ista*. What is striking in this concept is certainly its hierarchical order, which goes from the most general to the most specific term.

2. Groupements associatifs, microstructure (lexicale), sémantische Nische

Dokulil (1962: 75, 1986: 220) believes that WFT represents the most important WF characteristics and constitutes a nodal point in the nest of WF relations. A similar conclusion seems to emerge when looking at studies on WF in Latin. It is symptomatic that, especially from the beginning of the 1960s, an increasing number of publications pertaining to the classes of Latin derivatives with one specific affix have begun appearing, focused, from both synchronic and diachronic points of view, on both WF and semantic examinations of individual formations (Perrot 1962; Quellet 1969; Serbat 1975; Martínková 1980; Kircher-Durand 1983, 2002³; Fruyt 1986, 1989; Gaide 1988). It is worth noting, as well, that if the terminology concerning morphological structure of derivatives is usually, regardless of the linguistic current or national language tradition, rather well established – terms like *base*, *affix*, *suffix*, *prefix*, *formative* are widespread and well understood – the terminology concerning classes of words kept together by some common semantic and formal criteria is, in contrast, far from being unified. Linguists agree that the derivatives might be grouped together under these criteria, but, at the same time, different terms to name the distinguished items are used.

I was confronted with such terminological heterogeneity when I wanted to write in French about the classes of derivatives formed by specific suffixes in medieval Latin. Naturally, I had to familiarise myself with French terminology, first, which was, however, not always consistent with that of Dokulil (1962). The most systematic overview of the terms of WF as applied to Latin was given by Fruyt (1986). To define the items delimited by Dokulil as WFT, Fruyt uses the phrase “«groupements associatifs» ou «séries» des mots”,⁴ which she repeats after Marouzeau (1949). Basically, its contents are consistent with Dokulil’s definition, although it does not explicitly claim to be a unifying criterion of the words forming an associative group of their WF bases. In contrast, Fruyt emphasizes morphological, as well as phonetic, character of the bases in the context of the measurement of suffix

³ We mind tome IX of the series *Grammaire fondamentale du Latin, Création lexicale: la formation des noms par dérivation suffixale*, elaborated and edited by Chantal Kircher-Duran, which partially unifies selected chapters of the studies of different authors previously published elsewhere and delivers some completely new studies. In the Introduction, Kircher-Duran sketches methodological presumptions shared by the authors of the contributions, such as Saussure and Benveniste structuralism, Coseriu’s structural semantics, as well as more modern currents pointing out user’s linguistic sentiments.

⁴ “Il existe des «groupements associatifs» ou «séries» des mots ayant en commun un suffixe au niveau du signifiant, des traits sémantiques au niveau du signifié et des caractéristiques au niveau des référents” (Fruyt 1986: 23).

productivity, arguing that the more varying the bases to which a suffix may be attached, the more productive a suffix is and the wider its strength of extension should be viewed (1986: 23).

Fruyt (1986) lists, in addition, the terms used by other scholars to describe the groups of derivatives with common affixes being distinguished as somehow specific, which one can also consider as corresponding, more or less, to WFT in Dokulil's (1962) sense, namely: *microstructure*, *micro-structure lexical*, and *semantische Nische*. Dubois (2002), the author of the first term, interprets *microstructure*⁵ similarly to Dokulil's WFT, the class constituting subsystems within some larger structure, which, thanks to specific regularities, can function autonomously to some degree. Although Dubois does not at first explain what he means by specific regularities, his further explanations and especially the examples he provides – the same type of WF morphemes added before the same type of bases in French names of relatives – make it clear he thinks about the lexical items representing common semantics and, at the morphological level, common affixes.

The next phrase, *semantische Nische*, has been firmly accepted in German tradition. It was used for the first time by Baldinger (1950).⁶ Its definition has been recently developed by Hüning (2009: 183) who interprets *semantic niche* as the result of semantic fragmentation becoming apparent with the passage of time in word formation processes. Hüning defines it more precisely as “the group of words (subset of morphological category) kept together by formal and semantic criteria and extensible through analogy” (ibid.).⁷ Since Hüning employed it for contrastive examination (of the morphological and semantic differences between German and Dutch), he also outlines its usefulness for this kind of study. In his opinion, it allows for better generalizations about the systematic differences between two languages than the term *category*.

In turn, the phrase *micro-système lexical* has been consequently, and almost uniquely, used by Kircher-Durand (2002: 88, 149, 190, 198, 200). The context in which Kircher-Durand uses it shows that she means rather strict specific classes with this phrase. If one risked terminological comparison with Dokulil (1962, 1986), these should be placed somewhere between WFTs and WF subtypes. For example, in the case of the Latin suffix *-ensis*, primarily used to coin the names of inhabitants, Kircher-Durand delimits yet seven semantic fields in which it had been, starting from the Latin of the republic period, somehow productive. One of them is constituted by the epithets of deities derived from place names, and, as those establish a relatively extensive series, they are defined by her as *micro-système lexical* (2002: 191).

Incidentally, one yet encounters the phrase *sous-ensembles formeles*, also in the *sous-série* variant, introduced by Deaude (Kircher-Durand 2002: 290) concerning formations such as coined with *-ntia*, *-mōnia*, and *-itia* suffixes, which constitute a subset of the

⁵ [O]n donne le nom *microstructures* à certains sous-systèmes qui à l'intérieur d'une structure plus large, présentent des régularités spécifiques et une organisation qui leur assure une relative autonomie de fonctionnement. Ainsi, les noms des parenté constituent une *micro-structure formée*, en français, d'unités linguistiques en nombre fini déterminées sémantiquement par les rapports qu'elles entretiennent entre elles et par rapport à un moi (ego) imaginaire, et morphologiquement par un système particulier des morphèmes (grand et petit dans grand-mère, petit-fils, etc. (Dubois 2002: 304)

⁶ Metzler *Lexicon Sprache* (2010: 260) defines it as follows: “Ableitungen mit demselben Suffix bilden oft Reihen mit ident. WB [Wortbildungs]-Bedeutung, die K. Baldinger s. N. [semantische Nische] nannte”; “Derivatives with the same suffix often establish series with identical word-formation meaning which by K. Baldinger are named *semantic niches*”.

⁷ Hüning also briefly sketches the history of the term's interpretation (as by Erben 2000; Rainer 2003).

derivatives with the suffix *-ia*. Those terms seem to correspond to Dokulil's WF subtype more unequivocally.

When we compare the terms and phrases discussed, none of them so clearly emphasize the uniformity of the WF base like Dokulil. For Dokulil (1962: 75, 1986: 221), it is the foundation; for the derivatives forming common WFTs, besides the identity of general onomasiological and semantic structure, the identity of the affix must also represent the identity of the base (of the part of speech from which they are coined). Authors of other definitions speak about common morphological characteristics as well, but if any strict unity is explicitly postulated by them in this regard, it is that of the affix. However, they almost all underline that the discussed items constitute part of some larger structure ("sous-systèmes à l'intérieur d'une structure plus large" (Dubois 2002), "subset of morphological category" (Hüning 2009)). But again, none of them organize these classes in a hierarchically ordered framework to the same extent as Dokulil, which is reflected in the relevant terminology by Dokulil: *WF category* → *WFT* → *WF subtype*.⁸

However, returning to WFT and Latin, the phrase WFT as originally applied by Dokulil to synchronic analysis may also represent advantages for diachronic studies. In this perspective, it has been employed by Pultrová (2011), who accepted it as a classificatory framework when discussing the rules governing the rise, development, and differentiation of Latin formations established by the deverbative nouns and adjectives. It is interesting to note that Pultrová nowhere (in her monograph published in English) proposes any explicit definition of the term WFT⁹, instead, using it as if it could be understood by non-Czech readers. Pultrová apparently believes that it may be well understood through the context. That is perhaps the first time that the term WFT has been used on such a scale in an English linguistic text, i.e. used to repeatedly recall the principle of classification of lexical items, where, in addition, no initial extra explanation was given. Even if such a circumstance was caused by the fact that the discussed monograph is a translation from Czech, and for Czech readers such an explanation was not necessary, it shows, at the same, the extent to which the term WFT has been locally (in Middle Europe) accepted in the studies of WF. One can admit as well that such an approach reflects the author's conviction about the universal applicability of the term WFT, and that is, perhaps, the reason why Pultrová decided not to introduce it in a more explicit way.¹⁰

Pultrová points out advantages connected to the examination of historical linguistic processes within the structure of WFT. First, as it allows (2011: 12–14) proceeding from the function and not the form of the suffix, it helps to avoid interpretative problems resulting from functional and formal differentiation of the lexemes coined with the same suffix, in other words, from suffixal polysemy and a lack of uniformity in the base words of the distinguished lexemes. It is possible thanks to the perspective imposed by WFT that combines rigorously semantic characteristics of the items grouped together with their formal uniformity. Otherwise, no wonder that the author, who decided in her monograph to classify material according to WFTs, already in its title and once more in the introduction (2011: 15),

⁸ It is worth outlining that Dokulil's terminology had evolved with the passage of time. In Dokulil & Čechova (2011: 105 ff.), the term *category* is exclusively reserved for classifications at onomasiological level, while the previous *word-formation category* is defined as *word-formation class* (Czech *slovotvorná třída*). In addition, the phrase *word-formation subcategory* used in Dokulil (1962) is not encountered in Dokulil & Čechova (2011).

⁹ By Pultrová (2011) WFT = *word-formative type*.

¹⁰ In reality, the author, at the beginning, discusses generally Dokulil's concept of word formation and it is against this background that Pultrová introduces the term WFT (2011: 10–12).

precisely defines the categorical status of the base of the lexical items which Pultrová intends to discuss, namely that these will be deverbative nouns and adjectives.

Second, since the perspective of WFT is biased towards analysing the structure of the whole word, not only the suffix, it delivers the ground for examining the sound structure of lexical units within individual classes. Solving phonological problems from the point of view of WFT may significantly help the separation of inherited and analogical items, both in what concerns the distinguished items as a given WFT they enter and the individual formations (2011: 10).

The distinction of the inherited and analogical formations at more specific (within WFT) and more general (as WFT) levels, as can be observed in the study of Pultrová (2011), can be viewed as the third advantage owed to the application of WFT to diachronic studies.

Additionally, in the introductory chapter Pultrová (2011: 11–12) mentions the first, in effect, systematic classification of Latin derivatives, proposed by Leumann (1944); however, as she outlines, it failed to be put into practice. Leumann ordered his classification by the parts of speech, both of the word bases and the derivatives, and accordingly, thus delimited the main syntactic categories distinguished furthermore, as Pultrová names them, *traditional word-formative classes*, i.e. also taking into account semantic aspects, such as *nomina agentis* and *actionis*. Pultrová rightly points out the similarity of this concept to the classification put forward by Dokulil. As for terminological aspects, Leumann speaks about the kernel or main function (*Kernfunktion* or *Grundfunktion*) of the suffix, which should be distinguished, and stresses that this overlaps mostly with its historically primitive function.

3. Štichá's reinterpretation of Dokulil's framework

Dokulil's terminological apparatus has been recently newly approached by another Czech linguist, Štichá (2012, 2013, 2018), who tries to evaluate its usefulness especially from the point of view of WF analysis made with electronic corpora. Štichá has paid particular attention to the phrase WFT. On one hand, he points out (Štichá 2013: 341–344) its improper use (not consistent with Dokulil's definition), encountered by some scholars, but, on the other hand (2013: 342–344), its insufficiency for classifying data obtained from electronic corpora, insofar it is rigidly interpreted according to this definition. As a manifestation of improper use of the phrase, Štichá quotes relatively new studies, by Chýlová (2010) and Tůmová (2007), in which the WFTs were delimited derivatives not uniform with the part of speech of their bases, for instance, those derived from adjectives and those derived from substantives. Nevertheless, it should be emphasized that a less rigid interpretation of Dokulil's phrase has occurred before. For instance, Grzegorzczkova & Puzynina (1998: 384) claimed that within WFT one abstracts from the shape of the WF base (but, surprisingly, examples of the derivatives provided to illustrate their definition all represent uniformity of the base). Horecký (1999: 11, 2007: 26), in turn, interpreted the class of place names with the suffix *-áreň*, in Slovak, as one WFT that can make use both of a substantival and actional motivant.¹¹

¹¹ Horecký provides examples of desubstantive *lekáreň* 'pharmacy', *vináreň* 'wine cellar', as well as deverbal *sušiareň* 'drying room' and *plaváreň* 'swimming pool'. According to the terminology used by Dokulil (1962: 75), the classes characterized by the identity of the affix and semantics but not the base should be defined as WF subcategories.

Although Štichá believes (2018: 165) that the relation between the part of speech of the WF base and the affix of the derivative is crucial for WFT delimitation, he outlines, at the same time, that it would not be suitable to reduce such delimitations only to this relation. Even if it constitutes some general derivational model it, as he observes, does not allow considering, for example, the phenomenon of affixal polysemy. As he concludes, for the users of language, when choosing a suffix, the semantic category of a word being coined with it is often apparently not relevant. That is the reason why one, when delimiting WFT, should also consider categorical meanings of the derivatives. This consideration allows, then, distinguishing the categories in which a given suffix is preferably chosen by the users. The observation that polysemic affixes are in individual categories differentiated by their productivity is foundational for Štichá's interpretation of the term WFT.

It is also worth noting that the explicit definition of WFT that he proposes (Štichá 2018: 164) is rather general and similarly not only oriented on the relation between the base and affix. He states, "by word-formation type we understand, in general sense, the sum of characteristics of a given class of derivatives, insofar this class is defined by a given set of word-formation relations".¹² In addition, he recommends paying more attention to some of these relations, especially to the following: Is a derivation base constituted of a single or multi-word item? What operation was used to coin a new word (e.g. affixation, compounding)? Is the part of speech of the base word and the derivative the same or different? Are derivatives with the same affix formed from common base words (belonging to one given category) differentiated from the category they represent? As long as such semantic differentiation occurs and, in addition, a different degree of productivity between the derivatives in question can be observed, it may be, according to Štichá, an indication to distinguish different WFTs.

He points out, at the same time, the advantage delivered by big electronic corpora of easily and quickly discerning between frequent and not frequent classes (2013: 341). However, he doesn't postulate (2013: 342) some hazardous query but suggests identifying first, at least preliminarily, the typology of WF structures we want to examine in corpus. In fact, the purpose of such a search should be to objectify our presumptions, or, perchance, to discover some other regularities which we couldn't notice without examining a huge amount of texts. He even declares to be willing to consider WF categories and WFTs to be any category or type of structure being somehow specific (Štichá 2012: 97).

As one can see, in the delimiting of WFT, the central role is assigned to the productivity of word formation structures by Štichá. One can even risk the thesis that for him it becomes the main criterion of such delimitation. Consequently, Štichá rejects (2013: 344, 2018: 165) the phrase WF subtype as entailing the delimitation of too many specific structures, reserved by Dokulil for any more specific items within WFT. Štichá (2012: 98–99) accounts for Dokulil's classification framework (emphasizing the importance of WF subtypes) by the fact that he did not dispose of big text collections and while constructing his theory, mainly used data taken from dictionaries. So, he effectively could not have had the possibility to examine the productivity of the items he considered worth distinguishing as specific ones. However, it is worth underlining that Štichá appreciates Dokulil's intuition that for an exact WF analysis it is necessary to calculate productivity at different levels and degrees. As Štichá outlines (2012: 100–101), especially Dokulil's distinction between

¹² "Slovotvorným typem v obecném smyslu rozumíme souhrn vlastností určité třídy odvozených slov definované určitým souborem slovotvorných vztahů" (Štichá 2018: 164).

absolute and relative (depending on specific styles) frequency of WF methods, types and means can significantly contribute to the studies based on electronic corpora. Štichá (2018: 169 ff.), pursuing Dokulil's distinction of systematic and empiric, absolute and relative productivity proposes his own distinction, more appropriate as he believes, of systematic (potential or at the level of langue) and real (at the level of parole), as well as hidden productivity.

The real productivity can be established insofar as in linguistic and extra-linguistic circumstances it will be considered all together. Štichá stresses the importance that statistical examination of individual groups of derivatives aiming to distinguish the most common and rarest items has in this context. Štichá also postulates that, based on statistical examination, the degrees of productivity of the affixes operating within a given WFT should be determined (2018: 169–170). In fact, he not only determines such degrees (six in total), but also strictly defines the percentage of the derivatives coined with a given affix that should be decisive about enclosing this affix to one related degree or another.

Finally, according to Štichá (2018: 166), as an important criterion of the delimitation of WFT, insofar as one strictly combines it with productivity, the morphological structure of the base word and its derivational history (from which word and with what method it was coined) should be regarded. Štichá (2013: 344–346) also provides examples of WFTs as he delimited them, based on the data extracted from the Czech National Corpus, formed namely by the deadjective names of persons with the suffix *-ec*. He places here, indeed, particular emphasis on the influence which the morphological and phonological structure of the base words might have on the real productivity of the WFTs.

In contrast, a real polemic approach of Štichá towards Dokulil can be observed when the former states his opinion concerning onomasiological theory, which for the latter is essential in interpreting WF processes. Štichá redefines the position of WFT in Dokulil's terminological framework, subjecting two fundamental phrases in this framework to critical analysis, namely *onomasiological category* and *word-formation category*. He argues (2012: 97–98) that the very general character of onomasiological categories constitutes their strengths and, at the same time, their limitations and believes that those as superfluous might be well replaced just by *word-formation categories* – the term equally extensible in abstractness and generality.¹³ Insofar as WFT is concerned, he underlines that Dokulil's postulate of triple uniformity (of onomasiological category, WF base, and formative) entails difficulty when delimiting some WFTs because of the restraint number of basic conceptual categories admitted by Dokulil to represent the onomasiological structure. Since there are only four categories parallel to the main parts of speech (substance, action, quality, and concomitant circumstance) consequently, for example, the notion of *person* is not included into the set of onomasiological categories. And since for Dokulil the most important characteristic of WF base is the part of speech, consequently, for example, the Czech nouns *vyslanec* 'messenger' and *makovec* 'poppy-seed cake', both representing deadjective derivatives with the suffix *-ec*, are included in the same WFT by him. Štichá (2013: 344–345) strongly opposes such delimitations because, as he argues, the need to coin new names of persons usually results in language of a given period from other social conditions than the need to coin the names of non-persons. Thus, as he argues, the accuracy of the measurement of WFT productivity might be affected.

¹³ Štichá, in his most recent systematic presentation of the issue of WF and related terminology, in academic Czech grammar (2018: 163), declares he will not intentionally use either the phrase *onomasiological category* nor the adjective *onomasiological*.

4. WFT in Štekauer's onomasiological model

The phrase WFT is also employed by Štekauer as a part of the terminological system in his onomasiological model of WF. To better understand what place it occupies there, it is suitable to be reminded of some principles of Štekauer's concept. First of all, his interest is outside of the semasiological approach, which is effectively focused on analysing the results of the process of coining new words, on "the already existing word-stock" (Štekauer 2005: 207). The onomasiological method, having an interest in the naming act itself, focuses instead on the individual steps of the naming process (Panocová 2015: 45). According to Štekauer (2005: 226), the approach he proposes aims to interrelate the cognitive abilities of the speech community with both extra-linguistic and linguistic phenomena. Consequently, the level of traditional structuralist analysis focused on the examination of relations between formal and semantic characteristics of derivatives is in it degraded and subordinated to this conceptual one (cf. Štekauer 2005: 213, figure 1). Štekauer (2005: 226) wants to see his model as a reaction to the formalism of generative morphologists. For us, however, the more interesting question is how it is positioned with respect to the model of Dokulil. Štekauer (2005: 226) clearly exposes his objective, which is to pass over the borders between traditionally distinguished WF processes by accounting for them, no matter which WF method intervenes, "by means of the same word-formation principles." Such an approach manifests in combining the two above-mentioned points of view, namely the onomasiological and semasiological, in one framework. To be more accurate, Štekauer (2005: 215–217) distinguishes and includes into one framework onomasiological and onomatological levels of analysis, as the former is concept-grounded while the latter morpheme-grounded.

In contrast, when one reads Dokulil, one can get the impression that his presentation of onomasiological theory serves much more as a preparation for the subsequent exposition of WF rules.¹⁴ From this point of view, that he assigns such a high place in the nest of WF relations to the WFT should not be a surprise.

In turn, in Štekauer's model, the phrase WFT appears only when the computation of productivity is in question.¹⁵ In addition, it is rather not distinguished as an autonomous term, but as one of the terms of the four terms complex in relation, denoting four levels at which productivity is to be calculated (Štekauer 2005: 221 ff.). These are namely: onomasiological types, WF types, morphological types, and WF rules. At the first level, one examines the degree to which individual productive ways of forming naming units are involved. These ways may sometimes overlap with traditionally distinguished WF methods (like prefixation and affixation), but since the concept of onomasiological types is based on cognitive premises and starts from an extra-linguistic reality which is to be named, one abstracts from the WF (formal) perspective here. At this level, it is possible to evaluate the preference of language

¹⁴ Even if I am wrong in interpreting Dokulil's intentions, it is a fact that a similar approach, a virtual separation of the onomasiological level from the WF description, has been accepted as practice in some (if not in the majority of) studies referencing Dokulil. For example, in Polish grammar (Grzegorzczkova & Puzynina 1998: 390 ff.; Nagórko 2003: 214 ff.) authors do not discuss the principles of his onomasiological theory, but once noted that he departed from it and had delimited three categories of derivatives, namely mutational, modificational, and transpositional, they pass to their own classifications, based on the mentioned categories, without making further links, unless incidental, to the onomasiological theory. It should be added, however, that Grzegorzczkova devoted some lines to Dokulil's onomasiological concept in her previous, separate, volume (1973: 27) dealing with Polish word formation.

¹⁵ It should be underlined that in Štekauer's model, a foundational principle is that all WF and morphological types and rules are 100% productive and calculation is restricted only to actual words (Štekauer 2005: 222–223).

users for one of two contradictory tendencies: economy of speech and explicitness of expression.

At the second level, which is related to the conceptual categories (such as agent, location, and action), naming units (words) of different structure, defined just as different WFTs, are taken into account when computing productivity. Different structure means here a set of possible combinations of conceptual categories reflected in the structure of a given naming unit. If one took examples of these combinations provided by Štekauer (2005: 222) and applied traditional WF terminology to them, one would conclude that WFTs are delimited units regardless of the type of their base and the shape of their formatives; in other words, one abstracts here, once again, from a WF viewpoint. So, some substantial differences between Štekauer's and Dokulil's concept of WFT emerge here.

Štekauer proposes, in addition, the phrase *word-formation type cluster* as unifying various WFTs regarding the particular conceptual category, i.e. encompassing WFTs used "to coin new naming units falling within one and the same conceptual category". He admits, at the same time, the possibility that the productivity of an individual WFT within a given WFT cluster may be calculated. Due to the variety of representations of the WFTs at the morphological level, he delimits as the third level of productivity the calculation of morphological types. Similarly, the particular conceptual category distinguishes the morphological types cluster encompassing all morphological types used to coin new naming units within it. The individual morphological type may be calculated in similar terms to WFT clusters.

The fourth level, WF rules, is constituted by the unity of two precedents: WFTs and morphological types. Just this last phrase, *WF rules*, corresponds most closely to the Dokulil's WFT as it represents "the unity of the onomasiological and onomatological structures" (Štekauer 2005: 223). The examples of lexical units, provided by Štekauer, which might be put at this level, additionally strengthen this impression. Of course, the question remains whether this term has a chance to be employed as frequently as Dokulil's WFT. Although it has been regularly employed after 2005 (e.g. Körtvélyessy & Štekauer 2014: 413, 415, 419), the phrase *onomasiological type* seems to be of greater importance to the presentation of the onomasiological principles on Štekauer's concept.

To summarize our comparison, Dokulil and Štekauer use the term WFT in quite different senses; the latter entirely abstracts from a WF perspective. A more accurate equivalent to Dokulil's WFT is established by Štekauer's WF rules. These two phrases are not, however, of equal importance in the two respective models. While in Dokulil's model WFT represents a crucial point in the nest of WF relations, in Štekauer's WF rules intervene only at the end of the naming process, after the conceptual level is structured. An equal rank equivalent to WFT should rather be indicated by onomasiological types since as much in Dokulil's model lexical units are structured around WFTs as in Štekauer's one around onomasiological types.¹⁶

¹⁶ In practice, it can be observed when comparing the internal organization of the books or articles presenting respective points of view (as for Štekauer's model inspired studies, e.g. Panocová 2015: Chapter Five).

5. Lexico-semantic class vs. WF analysis and vs. automatic extraction of lexical data

The criterion of delimiting WFTs, as put forward by Štichá, based on productivity and emphasizing the role of big electronic corpora encourages, if not makes inevitable, methodological, as well as terminological, confrontation with the practice of classifying lexical items in the domain of computational linguistics. If we look for a term commonly used nowadays to classify sets of lexical items being extracted from big electronic corpora, it is, as it seems, the phrase *lexico-semantic class* (LSC), which merits particular attention. Obviously, it had been used before computational linguistics developed, and also in WF related studies.¹⁷ Among these studies, an important place is occupied by the monograph of Pounder dealing with the question of paradigms operating in WF morphology, since the framework of the derivatives classification by her adopted is actually based on LSCs, defined generally as “set of bases united semantically” (Pounder 2000: 663). However, it must be underlined that the distinction of individual LSCs is not the objective of this study, which is, indeed, the identification of class-specific paradigms productive in the particular LSC. Among those, Pounder understands small domains of systematic paradigms revealed by LSC as active (Pounder 2000: 662), which are, in turn, interpreted by Štekauer as a “subparadigm of systematic paradigm,” where the former differs from the latter only in size (Štekauer 2014b: 367–368). Since Pounder (2000: 663–664) wants to view LSC as “the main organizational principle of the lexicon for word-formation, within which recurring paradigm types play a role” one may define it, in a larger sense, as a semantic platform allowing the identification of further regularities, also morphological, of lexical items. One can, then, risk the hypothesis that from this point of view the function of LSC approaches somewhat like the WFT.

Pounder specifically examines German denominal and deadjectival adjective formations. An important conclusion Pounder comes to is that LSC allows the users of a language to make predictions relating to the set of the phenomena, mainly semantic, but also concerning form-meaning relations within the paradigm (2000: 662). Pounder wonders, furthermore, if the LSC may play a similar role in WF systems of other languages and if it has, in what concerns linguistic prediction, universal importance. Pounder provides a few examples from Russian and Labrador Inuktitut which make such presumption plausible (2000: 665). It is worth noting that, similarly, the concept of WFT stimulates Štichá to reflect on the possibilities of predicting lexical meaning based on the analysis of the WF structure of the derivatives (2018: 172–179). Of course, by Štichá the emphasis is more clearly laid at the starting point of the investigation, on morphology.

As stated, LSC as a phrase grouping lexical items has also occupied a particular place in the domain of computational lexical semantics. However, in this domain, the methods of its delimitation are usually abstracted from WF principles, emphasizing, by contrast, the role of context in which the words appear. One assumes (Jurafsky & Martin 2009: 671 ff.) that the examination of context similarity will enable one to detect semantic similarity. Retrieval of the LSCs is essentially based on the distributional properties of the word. The method of extracting LSCs may be either supervised or unsupervised – depending on whether we (or rather, the machine) dispose of hand-labelled (with correct sense) data or not when making a word sense disambiguation. In the latter case, the senses of a word are identified

¹⁷ We enumerate here some works in which LSC has been adopted as the framework for classification of lexical items, regardless of the degree to which the WF viewpoint is taken into account: Pounder 2000; Korhonen & Briscoe 2004; Gamallo et al. 2007; Romeo et al. 2012; Michow 2013; Malatowska-Statkiewicz 2014.

automatically, thanks to adequate machine learning procedure. When dealing with large corpora, unsupervised methods are preferred. They consist of building clusters both of words and local syntactic contexts. As a consequence, LSCs are generated through a clustering process. Each cluster, as the result of merging its most prototypical constituents (word and contexts), represents an LSC and is used as a so-called centroid to word classification. (Gamallo et al. 2007: 37 ff.)

Partington (1998: 138) defines LSC as “a set of lexical items which have a high probability of being found together in text or collection of texts of the same genre” and argues that statistically established co-occurrence of lexical items might reflect the way in which language users organize the world. Such a cognitively inspired presumption that the results obtained from the query of corpora are not random and might reveal certain regularities of which we are sometimes not aware is essential in computational linguistics. On the other hand, the methods of establishing LSCs based on automatic extraction of lexical data sometimes give the impression of having recourse to the criteria being too vague and abstracting too much from the morphological shape of the words. Partington enumerates an example of the lexemes forming one LSC with nouns such as *table*, *chair*, and *furniture* (1998: 138). I put aside, at this point, further discussion since it would digress from the paper’s central topic. My intention was here to show that the approach of computational linguistics may as much objectify the procedure of determining lexico-semantic regularities, as separate it from WF perspective.

6. Conclusion

The overview of the terms relating to the classification of lexical items being distinguished or emerging as somehow specific has shown that these terms, according to different methodological approaches, have served to bring out more or less different sets of common characteristics of concerned items. Special attention was given to the phrase WFT which was introduced by its author, Dokulil (1962), to individuate specific meeting points of morphological and semantic characteristics of the groups of derivatives. The phrase WFT appears, as well, in Štekauer’s (2005) terminological apparatus, where, nevertheless, it is involved only when the onomatological (morphologically oriented) level of the process of coining new lexical units is considered, according to the onomasiological interest of the author. Moreover, WFT does not occupy an equally important place as in Dokulil’s framework, in which it gets the status of the central point of WF relations; under Štekauer, WFT is reduced to the role of one of the components constituting WF rules, the phrase used by the author as a platform for unifying onomasiological (cognitively oriented) and onomatological levels. This last phrase is, effectively, that which corresponds most closely, insofar as such a comparison is justified, to WFT as understood by Dokulil.

In turn, what seems common to Dokulil and Štekauer is their ambition to propose a possibly holistic framework for describing WF relations. Their ambition is reflected in the hierarchically ordered notional systems they propose. However, while Dokulil, parting from onomasiological principles, focuses furthermore on the rules unifying morphological and semantic levels, Štekauer tends to liberate the interpretation of WF process from any formal restrictions, to examine it regardless of the WF methods and means actually chosen by the users of a language.

One has argued in this study that it is possible to indicate equivalents to the term WFT, as conceived by Dokulil, used by the scholars from other structuralist schools outside of Prague. However, the terms evoked in this context (*groupements associatifs*, *microstructure*, *semantische Nische*) do not seem to function, contrarily to Dokulil's work, as part of a wider, specially coined to describe WF relations and, in addition, hierarchically ordered apparatus. On the other hand, it is true that they almost all entail interpretations of lexical items being characterized by specific morphological and semantic identity as a subset of some more general structure.

One also paid attention to the possibilities which the concept of WFT may offer to strictly diachronic analysis of WF processes (Pultrová 2011). In the context of a diachronic approach, a possible utility of the phrase WF subtype was emphasized too.

In what concerns the development of the phrase WFT and, in a more general sense, of Dokulil's concept of WF, one gets the impression that they have undergone constant interpretation, not even as re-interpretation as much as deeper and more insightful interpretation, in the sectors already recognized by Dokulil as relevant for WF. The question of productivity stands out here as crucial. Dokulil's distinction of its different types and degrees has been resumed both by Štekauer and Štichá in the way to accept productivity as a kind of absolute criterion in their own concepts. Though Štekauer identifies WF with productivity and declares considering, in his model, only 100% productive WF rules, Štichá views productivity as reflecting virtual WF dynamics, creating the most objective criterion of delimitation of WFTs.

Since Štichá (2013) emphasizes the importance of big electronic corpora for the exact measurement of productivity, and postulates to admit them at large scale to the studies of word formation (what effectively already happens), a methodological and terminological confrontation between traditional and computational linguistics, will be, especially for the former, it seems, still a growing challenge. For example, the challenge of how the phrase LSC might be interpreted when being used for traditional WF analysis and how the automatic extraction of lexical items, as briefly shown throughout this study, remain. I am convinced that the phrase WFT imposes a certain, more systematic way of describing WF; in terms of considering regularities combining different linguistic levels and placing WF relations within some hierarchical framework, it may also successfully serve as the device of lexical classification when automatic lexical data extraction is in question. The first steps in this direction (Štichá 2013) are promising.

Abbreviations

LSC = Lexico-Semantic Class

WF = Word-Formation

WFT = Word-Formation Type

References

- Baldinger, Kurt. 1950. *Kollektivsuffixe und Kollektivbegriff*. Berlin: Akademie-Verlag.
- Chýlová, Helena. 2011. Odmítané slovotvorné typy v brusech jazyka českého. In *Minulost, přítomnost a budoucnost v jazyce a v literatuře* (Sborník z mezinárodní konference pořádané katedrou bohemistiky PF UJEP v Ústí nad Labem. II. díl.), 342–346. Ústí nad Labem: Univerzita J. E. Purkyně.
- Daude, Jean. 2002. Les substantifs abstrait de qualité. In Kircher-Durand, Chantal (ed.), *Grammaire fondamentale du Latin IX, Création lexicale: La formation des noms par dérivation suffixale*, 225–305. Louvain: Éditions Peeters.
- Dokulil, Miloš. 1962. *Tvoření slov v češtině 1. Teorie odvozování slov*. Praha: Nakladatelství Československé Akademie Věd.
- Dokulil, Miloš. 1986. Obecné pojetí tvoření slov. In Horálek, Karel & Knappová, Miloslava & Dokulil, Miloš & Petr, Jan (eds.), *Mluvnice češtiny 1: Fonetika, Fonologie, Morfonologie a morfemika, Tvoření slov*, 193–230. Praha: Academia.
- Dokulil, Miloš & Čechová, Marie. 2011. Tvoření slov. In Čechová, Marie & Dokulil, Miloš & Hlavsa, Zdeněk & Hrbáček, Josef & Hrušková, Zdeňka (eds.), *Čeština – řeč a jazyk*, 96–260. Praha: SPN - Státní pedagogické nakladatelství.
- Dubois, Jean & Giacomo, Mathée & Guespin, Louis & Marcellesi, Christiane & Marcellesi, Jean-Baptiste & Mével, Jean-Pierre. 2002. *Dictionnaire de linguistique*. Paris: Larousse Bordas VUEF.
- Erben, Johannes. 2000. *Einführung in die deutsche Wortbildungslehre* (Grundlagen der Germanistik (GrG), Band 17). Berlin: Erich Schmidt Verlag.
- Fruyt, Michèle. 1986. *Problèmes méthodologiques de dérivation à propos ds suffixes latin en -cus*. Paris: Klincksieck.
- Fruyt, Michèle. 1989. Etude sémantique des "diminutifs" latins: les suffixes -vlus, -culus, -ellus, -illus ...de-substantivaux et de-adjectivaux. Actes du Ve Colloque de Linguistique Latine (Cahiers de l'Institut de Linguistique de Louvain 15), 127–138. Louvain-la-Neuve: Peeters.
- Gaide, Françoise. 1988. Les «formes élargies» du «latin vulgaire»: un cas très particulier de la dérivation. *Latomus* 47(3). 584–592.
- Gaise, Françoise. 1988. *Les Substantifs masculins latins en -(i)ō, -(i)ōnis*. Louvain: Éditions Peeters.
- Gamallo, Pablo & Lopes, Gabriel P. & Agustini, Alexandre. 2007. Extraction of Lexico-Semantic Classes from Text. *PICS - Publication Series of the Institute of Cognitive Science* 1. 39–47.
- Glück, Helmut (ed.). 2010. *Metzler Lexikon Sprache*, 4th edition. Stuttgart: J. B. Metzler.

- Grzegorzczkova, Renata. 1973. Problemy kwantyfikacji w grupie werbalnej. In Topolińska, Zuzanna & Grochowski, Maciej (eds.), *Liczba, ilość, miara. Materiały Konferencji Naukowej w Jadwisinie, 11-13 maja 1972*, 83–99. Wrocław: Zakład Narodowy im. Ossolińskich, Wydawnictwo PAN.
- Grzegorzczkova, Renata & Puzynina, Jadwiga, 1998. Słowotwórstwo. In Grzegorzczkova, Renata & Laskowski, Roman & Wróbel, Henryk (eds.), *Gramatyka współczesnego języka polskiego. Moforlogia*. Warszawa: Wydawnictwo Naukowe PWN.
- Horecký, Ján. 1999. Onomaziologická interpretácia tvorenia slov. *Slovo a slovesnost* 60(1). 6–12.
- Horecký, Ján. 2007. An onomasiological interpretation of word-formation. *SKASE Journal of Theoretical Linguistics* 4(1). 21–27.
- Hüning, Matthias. 2009. Semantic niches and analogy in word formation. *Languages in Contrast*, 9(2). 183–201.
- Jurafsky, Daniel & Martin, James H. 2009. *Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition. Upper Saddle River, NJ: Pearson Prentice Hall.
- Kircher-Durand, Chantal. 1983. Les noms latins en -nus, -na, -num. Étude morpho-sémantique et historique d'une catégorie dérivationnelle du latin classique. *L'Information Grammaticale* 16, 40–52.
- Kircher-Durand, Chantal (ed.). 2002. *Grammaire fondamentale du Latin IX, Création lexicale: la formation des noms par dérivation suffixale*. Louvain: Éditions Peeters.
- Korhonen, Anna & Briscoe, Ted. 2004. Extended Lexical-Semantic Classification of English Verbs. *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*. Boston: Massachusetts.
- Körtvélyessy, Livia & Štekauer, Pavol. 2014. Derivation in a social context. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 407–423. Oxford: Oxford University Press
- Leumann, Manu. 1944. Gruppierung und Funktionen der Wortbildungssuffixe des Lateins. *Museum Helveticum* 1(3). 129–151.
- Malanowska-Statkiewicz, Maria. 2014. Nazwy pokrewieństwa i powinowactwa w języku czeskim i polskim. Paradygmat słowotwórczy. *Slavica Wratislaviensia* 159. 245–254.
- Martínková, Dana. 1980. Slova tvořená sufixem -ista ve slovní zásobě středověké latiny v českých zemích. *Listy filologické* 103. 144–147.
- Michow, Elżbieta. 2013. *Studia nad frazeologią somatyczną języka polskiego i bułgarskiego*. Poland: Kielce.
- Marouzeau, Jules. 1949. *Quelques aspects de la formation du latin littéraire*. Collection linguistique (Société de Linguistique de Paris) 53. Paris: Klincksieck.

- Nagórko, Alicja. 2003. *Zarys gramatyki polskiej*. Warszawa: Wydawnictwo Naukowe PWN.
- Panocová, Renáta. 2015. *Categories of Word Formation and borrowing. An Onomasiological Account of Neoclassical Formations*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins Publishing.
- Perrot, Jean. 1962. *Les dérivés latin en -men et -mentum*. Paris: Klincksieck.
- Pounder, Amanda. 2000. *Processes and Paradigms in Word-Formation Morphology*. Berlin: Mouton de Gruyter.
- Pultrová, Lucie. 2011. *The Latin Deverbative Nouns and Adjectives*. Prague: Charles University, Karolinum Press.
- Quellet, Henri. 1969. *Les dérivés latin en -or*. Paris: Klincksieck.
- Rainer, Franz. 2003. Semantic fragmentation in word-formation: The case of Spanish -azo. In Sinch, Rajendra & Starosta, Stanley (eds.), *Explorations in seamless morphology*, 197–211. New Delhi: Sage Publications.
- Romeo, Lauren & Mendes, Sara & Bel, Núria. 2012. Using qualia information to identify lexical semantic classes in an unsupervised clustering task. *Proceedings of the 24th International Conference on Computational Linguistics COLING, 8–15 December, 2012; Posters*, 1029–1038. Mumbai: The COLING 2012 Organizing Committee.
- Rzepliela Michał. 2005. Les néologismes créés par suffixation dans le latin polonais. *Archivum Latinitatis Medii Aevii* 63. 35–44.
- Serbat, Guy. 1975. *Les Dérivés nominaux latins à suffixe médiatif*. Paris: Klincksieck.
- Štekauer, Pavol. 2005. Onomasiological Approach to Word-formation. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-formation*, 207–232. Dordrecht: Springer.
- Štekauer, Pavol. 2014a. Derivational Paradigms. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of Derivational Morphology*, 354–369. Oxford: Oxford University Press.
- Štekauer, Pavol. 2014b. Derivation in a social context. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 407–423. Oxford: Oxford University Press.
- Štichá, František. 2012. Jak v epoše elektronických korpusů následovat Miloše Dokulila [How to follow Miloš Dokulil in the era of corpora] (Miloši Dokulilovi ke stému výročí narození). *Jazykovědné aktuality* 49(3/4). 95–107.
- Štichá, František. 2013. K pojmu “slovotvorný typ” [Concerning the concept “derivational type”]. *Philologica* 72. 341–347.
- Štichá, František. 2018. *Velká akademická gramatika spisovné češtiny. I. Morfologie: Druhy slov / Tvoření slov. Část 1*, 162–179. Praha: Academia.

Tůmová, Šárka. 2007. Jeden slovtvorný typ v současné němčině (Ze sborníku Zur Rolle der Sprache im Wandel der Gesellschaft/The Role of Language in Changes of Society, Turku 2002). *Časopis pro moderní filologii* 89(1). 55–59.

Michał Rzepiela
Instytut Języka Polskiego Polskiej Akademii Nauk
al. Mickiewicza 31, 31-120 Kraków
michal.rzepiela@ijp.pan.pl

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/10.pdf. ISSN 1336-782X.