The schematization of Hungarian participle-noun compounds

Nóra Kugler
Eötvös Loránd University

Gábor Simon
Eötvös Loránd University

Abstract

In this paper, we present a new approach to compounds, arguing that compositional descriptions do not efficiently account for their formation, and emergence. We interpret the emergence of participle-noun (PTCP-N) compounds from phrases as guided by a complex schematization process which makes varied instantiations possible and results in the emergence of schematic constructions. The phenomenon's variability is demonstrated by a corpus study of Hungarian compounds with *érintő*- ('touch.PTCP, touching'), whereas the role of schematization is confirmed by the findings of a questionnaire study. By subjecting our corpus data to both quantitative and qualitative analysis (relying on Cognitive Grammar), we describe the rich variability of *érintő*-N compounds. In addition, a questionnaire study was used to confirm the emergence of schematic constructions, their entrenchment and the hypothesis of passive construal being a key motivating factor behind compounding. Employing the control cycle model of Cognitive Grammar, we propose a comprehensive account for the emergence of compounds with a PTCP initial component. Whichever participant is profiled, the cycle is always implemented in the case of compounds with an initial PTCP. Thus, this model describes the emergence of compounds as a process in which semantic integration between the components occurs as part of a more abstract process of schematization.

Keywords: participle-noun compounds, schematization, emergence, empirical methods, control cycle

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Introduction

In this paper we present a novel, meaning-centred account of compounding which employs the methods of cognitive linguistics and corpus linguistics. We propose a new model for the semantic description of compounds, which also implies a different view of how compounds emerge than previous accounts. The key thesis of the paper is that compounding is motivated by an increase in semantic integration between the components. Concomitantly, reconfigurations in the components' meanings produce higher conceptual proximity. At the same time, varied, repeated instantiations of tight semantic integration, motivated by the control cycle (cf. Langacker 2009: 130–135; 2016), also result not only in particular compounding patterns of individual expressions but also in the abstraction of new constructional schemas of compounding, which incorporate the emergent meaning of compounds in a more schematic fashion. As a consequence, the schematic constructional meaning serves as a new baseline for creating novel compounds. The increase of semantic integration (and conceptual coherence) between the components is accompanied by the schematization of component meanings, thus the specific constructional schemas emerging from the control cycle make compound formation productive to a certain degree. The detailed analysis of corpus data can provide a comprehensive approach to the productivity of compounding, as well as to the patterns observed in language use.[1]

The validity of these theses is demonstrated in this paper by a methodologically complex, extensive corpus study, which also anticipates a more comprehensive description of Hungarian compounds whose first component is an *-Ó* participle (PTCP).[2] With the aim of producing an in-depth analysis, we focus on one specific participle component, namely *érintő* 'touch.PTCP, touching', occurring in compounds such as *érintőceruza*

[1] We use the term schema as a motivating structure (with unipolar or bipolar organization) which is part of the knowledge of the language user. Construction is a more elaborated symbolic structure (with a phonological and a semantic pole), a specific type of a schema. Finally, by pattern we mean complex linguistic expressions occurring in a corpus or in the data of the questionnaire.

[2] In Hungarian there are three morphemes indicating participle: *-Ó* (*-ó*/*-ő*), *-T* [*-t*/-(VOC)*tt*] and *-AndÓ* (*-andó*/*-endő*). The difference between the three participle affixes is that in the temporal relationship between the foregrounded process (indicated by the verb), and the backgrounded process (indicated by the participle), and in force dynamics (the *-AndÓ* participle construes the process as which should be done, therefore opens the semantic domain of modality).

'touch.PTCP-pencil', *érintőképernyő* 'touch.PTCP-screen' and *érintő-alkalmazás* 'touch.PTCP-app'. However, the construction grammatical description of the schematized structure (and its constructionalization) is beyond the scope of the present paper.

Our research relies heavily on previous results on compounds obtained in traditional or cognitive linguistic frameworks; however, it is still novel in many respects. In line with the cognitive perspective, we focus on meaning rather than syntactic structure. Since PTCP-N compounds follow the modifier-head structure, i.e. they can be analysed as nominal compounds (Dirven & Verspoor 2004), the literature mostly approaches linguistic data from the perspective of the head component. However, from a cognitive grammatical point of view, the head noun specifies a schematic substructure of the modifier in the course of semantic integration, thus semantically the noun depends on the modifier. Therefore, our priority will be the semantic analysis of the composite structure rather than syntactic constituency. Although we make no attempt in this paper to provide a comprehensive treatment of all PTCP-N compounds in Hungarian, at the present stage of the research it seems that the varied types of head noun correlate strongly with the processual meanings symbolized by the PTCP component. Consequently, conventional syntactic analyses of compounding in Hungarian (see Kiefer 2000) may at best serve as a point of departure, with semantic accounts offering a fundamentally different picture.

The present research is based on the complex theoretical framework of cognitive linguistics (including cognitive grammar, and cognitive corpus linguistics). However, compared to previous cognitive linguistic treatments of compounding, our approach to the phenomenon is different in some ways. While we regard conceptual proximity of the component meanings as an important motivating factor, we are primarily looking at linguistic structures rather than conceptual representations. Cognitive semantic analyses generally present the emergent meaning of compounds by reference to conceptual integration (see Dirven & Verspoor 2004: 55; Benczes 2006; Ungerer 2007: 655–656); however, both Bundgaard et al. (2006) and Heyvaert (2012) express reservations. Here, we intend to offset the shortcomings of conceptual descriptions by a detailed cognitive grammatical analysis of semantic integration. Moreover, we explore constructional schemas which motivate the formation of individual compound expressions. Our explanation of schematization is based on the notion of control cycle (Langacker 2009: 130–135; 2016). The control cycle is a general model of semantic change in which an entity (the target) becomes an integrated

subpart of the conceptual domain of an actor as it is brought under the actor's control (see §6.2 for more detailed description of the process). As it is demonstrated below, the target entity is symbolized by the nominal component of the PTCP-N construction, whereas the verb stem of the participle expresses the actor's control. The control cycle serves as a basic cognitive model in explaining the conceptual organization of compounds. In addition, the model helps us to refine and specify that "micro-narrative scenario" (or "teleological frame of purpose-related action"), which – according to the proposal of Bundgaard and his colleagues – serves as "a 'schematic algorithm' underlying the meaning-construal of compounds" (Bundgaard et al. 2006: 375).

Another recurring point in the literature on compounding is the existence of an endocentric/exocentric continuum (Benczes 2006: 8–9). The traditional definition of endo- and exocentric compounds rests on the notion of head: endocentric compounds have a head, while exocentric compounds have no head (Scalise et al. 2009: 49). From a cognitive grammatical perspective, in the case of endocentricity one of the components serves as profile determinant for the whole; exocentricity means here that none of the components behave as profile determinant. From this point of view, our analysis will be novel by mapping the scale frequently mentioned in the literature via the corpus study of a single component's patterns of use rather than a lexically more heterogeneous sample of data. We argue that data which can be assigned to the exocentric domain are the products not only of creative innovation (cf. Benczes 2006), but also of a regular process (i.e. motivated by constructional schemas) of language change (see also Scalise et al. 2009 for typological data).

The most common approach to compounding is based on the principle of compositionality. Even the cognitive linguistic analysis of the N+N construction by Benczes (2006) aims to establish which component functions as head or modifier, and how they contribute to the meaning of the compound as a whole (this serves as a basis for arranging data on the scale of exocentricity). In other words, previous studies are characterized by (more or less) latent compositionality (see also Bundgaard et al. 2006: 366–368). One of the most important consequences of this theoretical orientation is that previous models can deal only with individual data and not with a complex pattern of the same substructure having different meanings. By contrast, the model we are proposing replaces compositionality by the baseline/elaboration relation hinging on the control cycle, and explains the formation of compounds by the development of a more abstract, emergent

meaning rather than by the concatenation of component elements. In our view, compounding involves a reorganisation of the semantic structure of the components' meanings on a higher level of abstraction, which foregrounds novel aspects of the components' meanings.

Methodologically, our research is novel by demonstrating scalarity not through psycholinguistic experiments (cf. Ryder 1994; Benczes 2006; Ungerer 2007) but rather by quantitative, corpus linguistic analysis of a large body of data. Moreover, the data collection method of questionnaires (meaning attribution to nonsensical expressions) was targeted not at the identification of meanings or at the elicitation of acceptability judgments but rather at the demonstration of schematization. In addition to the cognitive grammatical explanation, we implemented the corpus-based measurement of the components' semantic accommodation, a characteristic property of compounds.

The paper first presents the central questions and hypotheses behind our research (§2). This is followed by the presentation of methods and the material under study (§3). We first offer a detailed overview of categories identified by corpus data (§4), then discuss the data gained by questionnaire (§5), and compare the results obtained from the two sources (§6), before supplementing them with an account drawing on the control cycle as interpreted in Langacker's Cognitive Grammar (§6.2). Finally, the last section offers a short summary and concluding remarks (§7).

2 Research questions and hypotheses

At the centre of our research are compounds as symbolic units. In terms of Cognitive Grammar (Langacker 2008), the semantic pole of a compound is an integrated meaning in which the components constitute a coherent and tight conceptual unit, and this Gestalt character is iconically expressed on the phonological pole by single-word spelling and a reconfiguration in stress patterns (as compared to phrases). Therefore it is important to emphasize that the structure results from the emergence of a schematic meaning allowing for conceptual unity. It is significant, moreover, that the emergence of a new structure is motivated by changes on the semantic pole, with formal devices (including two-word, hyphenated and one-word spellings, and the instantiation of various stress patterns) merely symbolizing different degrees of semantic schematization. Thus, the main aim of the research is to explore the process of semantic change or reorganisation (schematization) by closely

examining a specific compound type (PTCP-N) as well as a specific usage (the *érintő* 'touch.PTCP, touching' – N) in Hungarian.

From these background assumptions it follows that our research focuses on the following questions. Firstly, how is the conceptual relation inherent in PTCP-N compounds to be characterized, which allows for the integration of components? Secondly, what kind of increasingly schematized meaning makes the constructions under study conventional, and how is this schema to be explored? Do empirical data support a specific model of schematization?

Bundgaard et al. (2006: 383–386) point out that the construal of a compound's meaning relies on schematic event frames. One of them is the agent → act → instrument → object/patient → result/goal frame, and the meaning of a compound depends on what is the focus (or window of attention) in the frame, i.e. which aspect of the frame becomes foregrounded by the nominal component. Relying on this, when an *érintő*-N compound highlights the agent (or metonymically the instrument, see §3.2 for more details) of the process of TOUCHING, we consider it to have active meaning. By contrast, in the case of foregrounding the object/patient or the result of the process, the meaning of the compound is considered passive. Evaluation in the active/passive dimension is of course a matter of degree rather than a clear-cut distinction.

Next let us turn to the hypotheses. Whereas phrasal structures have a basically active meaning,[3] i.e. the head noun profiles the agent of the process denoted by the participial attribute (e.g. *a csövet*$_{\text{PAT}}$ *ásóval*$_{\text{INSTR}}$ *érintő munkás* → the pipe.ACC$_{\text{PAT}}$ spade.with$_{\text{INS}}$ touch.PTCP worker.NOM 'the worker touching the pipe with a spade'), the nominal component of compounds profiles some other (non-agent or less agentive) participant, a means or patient of the action[4] (e.g. *érintőceruza*$_{\text{INSTR}}$ 'touch. PTCP-pencil', *érintőképernyő*$_{\text{PAT}}$ 'touch.PTCP-screen'). We hypothesize that the conceptual

[3] Note, that Hungarian PTCP's of the form V-*ó/-ő* typically have an active meaning, with passive interpretation counting as exceptional in this morphological construction (an example being *eladó lakás* 'away$_{\text{PREF}}$-sell.PTCP flat; flat to be sold', lit. 'selling flat'). However, morphological constructions of the V-*hAt-Ó* structure usually have a passive meaning (e.g. *eladható lakás* 'away$_{\text{PREF}}$-sell.POT.PTCP flat; flat that can be sold'), and it is the active interpretation that is exceptional (*a döntőbe bejutható* [in$_{\text{PREF}}$-get.POT.PTCP] *versenyzők* 'competitors who can get into the final'). Hence, the derivational suffix of PTCP forms can only be given construction-specific characterizations with respect to the active/passive dimension.

[4] Active or passive construal interacts with the grammatical voice of the verb stem, but it is not identical with it.

motivation behind compounding lies in the expression of passivized meaning. However, this change is a gradual process, following successive phases of the control cycle, therefore we expect that the phenomenon cannot be reduced to binary oppositions (as two-word vs. one-word spelling might suggest on the formal side) but rather there is a range of intermediate meanings. Particular schemas may thus be arranged with respect to each other.

3 Material and methods

3.1 Outline of the research

We investigated the emergence of compounds and the corresponding schematization of meaning with various tools and methods. First we tested the validity of our hypotheses on corpus examples of a more specific Hungarian PTCP-N structure, namely *érintő*-N (touch.PTCP 'touching'-N). In our analysis, we chose compounds with *érintő* 'touching' as their first component because we had experienced the high frequency of *érintőképernyő* 'touch.PTCP-screen' in colloquial speech, and here the participle occurs as a component of a passive composite structure. Investigating the corpus data made it possible to analyse certain patterns but did not support directly any conclusions about the emergence of schematic meaning of Hungarian PTCP-N construction. Therefore, we also elicited data by a questionnaire which explored how PTCP-N structures are processed in general.

3.2 The complex meaning of *érintő*-N

It is worth starting the analysis with a Cognitive Grammar account of phrases in which the head noun is modified by a participial attribute, since this structure is the point of departure (or baseline stratum, cf. Langacker 2016) of the hypothesized schematization process. In addition, some general aspects of the meaning of the *-Ó* participle must be clarified before describing how it functions in compounds. In the Hungarian phrase *a csövet*$_{\text{PAT}}$ *ásóval*$_{\text{INSTR}}$ *érintő munkás* (the pipe.ACC$_{\text{PAT}}$ spade.INS$_{\text{INSTR}}$ touch.PTCP worker.NOM$_{\text{AG}}$; 'the worker touching the pipe with a spade'), the processual meaning of the participle is fully instantiated, with the head noun profiling the primary, agent participant of the process (while the secondary participant functions as patient, and the instrument is also expressed). In *a*

kapcsolót$_{\text{PAT}}$ *érintő ujj*$_{\text{INSTR[→AG]}}$ (the switch.ACC$_{\text{PAT}}$ touch.PTCP finger.NOM$_{\text{INSTR[→AG]}}$; 'the finger touching the switch'), which is also a noun phrase rather than a compound in Hungarian, the processual meaning is again manifested; however, the component profiling the instrument is construed as primary participant. At the same time, the conceptualizer extends to it the agentive role on a metonymical basis. The extension is metonymical, since in Hungarian, *ujj* 'finger' can be the subject of the verb *érint* 'touch', and can be understood as the most active participant in the immediate scope, although the maximal scope of the verb's meaning involves the causer of the process. The secondary participant is again the patient here (Figure 1).

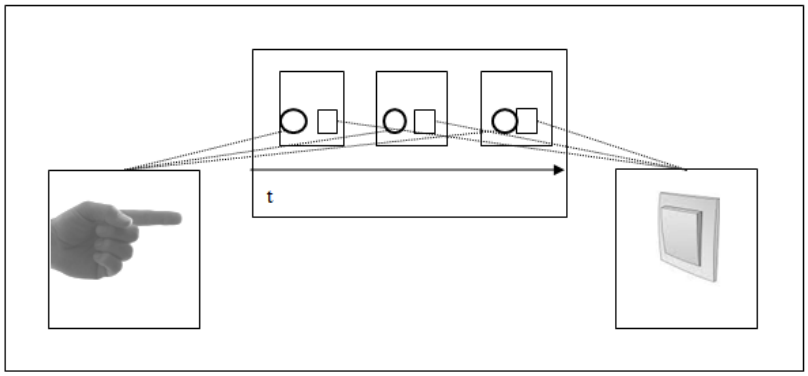

Figure 1: Semantic integration in the semantic structure of *a kapcsolót érintő ujj* 'the finger touching the switch'

These examples suggest that the meaning of phrases with a participial attribute (*érintő* 'touch.PTCP; touching' N) is primarily active (the head noun is the agent of the process symbolized by the participle) and secondarily active via metonymic extension (the head noun is construed metonymically as agent of the process). Thereby, the meaning of the structure undergoes a slight shift from active toward less active meaning (as the primary participant can be an agentive instrument), already in the case of phrasal constructions.

The compound (*érintő*-N) also evokes the process and its participants but in a different way. In the course of semantic integration, only a certain phase of the process (the moment of touching, e.g. in *érintőceruza* 'touch.PTCP-pencil') can be conceptualized, or else the process as a whole in a highly schematic fashion (e.g. *érintőszoftver* 'touch.PTCP-software'). In addition, several compounds direct attention at the secondary participant (e.g. *érintőképernyő* 'touch.PTCP-screen'). Thus, the metonymization process observed with attribute + noun constructs is in full swing here, and

the patterns resulting from it profile the passive (patient) participant or result of the process rather than its active participant. This is called passivized meaning in the present paper.

3.3 Corpus-based analysis

We studied a total of 6218 compounds beginning with *érintő-* 'touching', retrieved from the Hungarian National Corpus[5] (HNC, cf. Oravecz et al. 2014), which as a database meets the requirements of coverage and complexity.[6] With regard to the searching and filtering of results, the following remarks are in order. We performed a lemmatized search for compounds, which allowed us to retrieve all the relevant suffixed forms from the database. We ignored words and expressions involving the mathematical meaning of *érintő-*, namely 'tangent' (e.g. *érintőirány* 'tangent-direction', *érintőnégyszög* 'tangent-quadrilateral', *érintőút* 'tangent-road', *érintőlövés* 'tangent-shot') and other technical terms (e.g. *érintőhang* 'ballistic consonant, tap' in phonetics), since as lexicalized items they are not relevant for a study of the continuum between phrases and compounds. Such data (a total of 98 tokens) were removed from our sample.

Our full sample included hyphenated forms (*érintő-képernyő* 'touch.PTCP-screen') and multiple compounds as well (*érintőszenzor-rendszer* 'touch.PTCP sensor system', *érintőkijelző-barát* 'touch.PTCP-screen friendly'). Our procedure for hyphenated forms was as follows: (i) two-component *érintő*-N compounds (7 tokens) were not subjected to quantitative study, but their qualitative analysis was performed; (ii) multiple compounds involving *érintő*-N (938 tokens) were also included in the sample for quantitative analysis, since in these cases the hyphen connects an additional component to the construction, which suggests that the *érintő*-N pattern functions as a compound by itself (in contrast with *érintő-multimédiás* [N] 'touch.PTCP-multimedial [N]', in which *érintő-* 'touch.PTCP; touching' can be interpreted either as component of a compound or as part of a phrase).

[5] http://corpus.nytud.hu/mnsz/index_eng.html (Accessed 2018-03-16.)

[6] The new version of Hungarian National Corpus is the most comprehensive source for investigating Hungarian with corpus linguistic methods: beside the samples from written texts (having been available already in the old version of the corpus) it also includes data from spoken Hungarian and from social media as well (Oravecz et al. 2014: 1719). Thus it represents the rich variability of the use of Hungarian with regard to both topics and genres.

3.4 Questionnaire

Our online questionnaire had informants attribute meaning to eight nonsensical expressions (*csatit rimpít, tirimpon zargál, rimpuló tami, csatipivogoló, csatipivogolás, zargált miró, pivigolótami, parilva rilmütyöl*).[7] The instruction was given in the following form: "Please provide the meaning of the expression by using the words involved. Example: *pricsiges paró* 'a paró characterized as having pricsig (e.g. it has a pricsig in it, on it, or it possesses a pricsig)'."

The data directly relevant for our research are informants' analyses of *pivogolótami*, since this expression follows the schematic phonological structure of PTCP-N compounds and hence it can be interpreted as a PTCP-N compound as well. In order to find out about the relationship between phrases and compounds, we further compared *rimpuló tami* and *pivogolótami* (note that this analytic goal had motivated the inclusion of the same nonsensical component in both).

The questionnaire was filled in by 50 participants, 41 women and 9 men, aged 28.6 on average (with the youngest informant at 18 and the oldest at 61 years of age). 83% of our informants had a university degree. We did not study correlations between responses and sociolinguistic variables, thus no further data about informants is provided here.

4 The data of the corpus

4.1 Aspects of categorization

The key criterion of our analysis of corpus tokens is which participant of the processual meaning of the participle is profiled by the head noun, and how the participant thus profiled is construed against the conceptual base of the entire process.[8] We use the analytic categories of Cognitive Grammar to

[7] The expressions in the questionnaire includes some parts that can be identified as morphemes or morpheme structures (e.g. a word ending with *-t* (in *csatit*) can be interpreted as accusative form of a noun; or ending with *-oló, -ólás* (in *pivogoló, pivogolás*) can be understood as deverbal derivative forms). However, the nonsensical expressions themselves did not initiate symbolic interpretation (i.e. they are unipolar structures since the basis of their interpretation is purely phonological in this case, see Langacker 2008: 174).

[8] Bundgaard et al. (2006: 383) formulate this aspect of analysing compounds in almost the same way, as follows: "the *XY profiles* aspects of the event, whereas the event frame (contributed by *Y*) as such is the *base* for this profile."

refer to the schematic meanings of particular compounds. Our background assumptions suggest that passivization is not a homogeneous process but rather it results in various constructional schemas, making alternate construals available. But the relationship between these schemas is not derivational in character, therefore there is a range of alternative schemas available for the language users at a given time. Accordingly, the corpus data are highly uneven, with categories varying in their frequencies of instantiation. In addition to specifying token frequency, we also noted proportions within a given category in order to assess analogical productivity. In the sections below, categories of corpus data are discussed in turn.

4.2 The profile is the primary figure as INSTR

Here we included structures in which the noun component expresses the agent of the process (via metonymic shift of attention, with the immediate scope restricted to the participant performing the action and excluding the causer). Thus, the compound refers to a thing which is 'used for touching (other objects)' or 'designed for touching', e.g. *érintőtoll* 'touch.PTCP-pen' as a compound profiles the device with which touching is accomplished, in other words, it foregrounds the instrument of touching (see Table 1, $p = 44$, $n = 6218$).

Table 1: Frequency data on the 'instrument' category (C1) in the corpus

	Number of tokens	**Proportion within category (%)**
érintőceruza 'touch.PTCP-pencil'	38	86.4
érintőpálcika 'touch.PTCP-stick'[9]	2	4.5
érintőtoll 'touch.PTCP-pen'	2	4.5
érintőhőmérő 'touch.PTCP-thermometer'[10]	1	2.3
érintőpálca 'touch.PTCP-stick'	1	2.3

Within the full sample, tokens belonging to the first category only account for 0.7% (see Table 6 below for the details). At the same time, note that this composite structure comes closest in its meaning to the phrase chosen as our point of departure (*érintő ujj* 'touch.PTCP finger$_{\text{INSTR}[\rightarrow\text{AG}]}$, touching finger'), which already profiles as primary figure the instrument also functioning as the performer of caused motion (by way of a metonymic shift of attention). The compound functions in almost the same way as the phrase (which may explain its low token frequency); however, performing the process expressed by the participle is not construed as a one-off occurrence but rather is something that the entity is meant to be doing, which provides sufficient motivation for compounding and one-word spelling.

4.3 The profile is the secondary figure as touched surface

Here we included compounds in which the noun elaborates the secondary figure of the process designated by the participle, i.e. the schematic entity 'that can be touched'. For example *érintőfelület* 'touch.PTCP -surface' refers to a surface that requires touching during its application (cf. *érintőtábla* 'touch.PTCP-board'). Further, these expressions have it in common that they construe the processual meaning of the participle as a prototypical physical process, whose landmark (the touched entity) is consequently elaborated as a physical entity or more particularly its surface (see Table 2, $p = 6117$, $n = 6218$).

[9] The noun *pálcika* is a diminutive derivational form of *pálca* 'stick'.

[10] For the sake of clearness, it is worth adding that in contrast with non-contact types, touch-thermometer measures the temperature of the body through physical contact with its surface. The device needs a contact with the human body in order to register its temperature, but the functioning of the thermometer is not the result of the touching process on its own.

Table 2: Frequency data on the 'surface secondary figure' category (C2) in the corpus

	Number of tokens	**Proportion within category (%)**
érintőképernyő 'touch.PTCP-screen'	3163	51.71
érintőkijelző 'touch.PTCP-display'	2374	38.81
érintőgomb(*sor*) 'touch.PTCP-button(row)'	264	4.32
érintőpad 'touch.PTCP-pad'	84	1.37
érintőfelület 'touch.PTCP-surface'	75	1.23
érintőpanel 'touch.PTCP-panel'	58	1.00
érintőbillentyű(*zet*) 'touch.PTCP-key(board)'	27	0.44
érintőszenzor 'touch.PTCP-sensor'	24	0.39
érintőlap 'touch.PTCP-flat' ('touchpad')	23	0.38
érintőakna 'touch.PTCP-pit'	7	0.11
érintőtábla 'touch.PTCP-board'	4	0.07
érintőfólia 'touch.PTCP-foil'	3	0.05
érintőmonitor 'touch-PTCP-monitor'	3	0.05
érintőkapcsoló 'touch.PTCP-switch'	2	0.03
érintőegér 'touch.PTCP-mouse'	1	0.02
érintőjelző 'touch-PTCP-indicator'	1	0.02
érintőmegjelenítő 'touch.PTCP-display'	1	0.02
érintőgép 'touch.PTCP-machine' ('touch-pad')	1	0.02
érintőcsap 'touch.PTCP-tap'	1	0.02
érintőgitár 'touch.PTCP-guitar'	1	0.02

This category proved to predominate in terms of both type and token frequency (see Table 6). In the discussion of results of the corpus study, it will receive a more detailed treatment (see §4.7).

4.4 The profile is a virtual secondary figure

The compounds in this category profile a virtual space as the secondary figure of the process symbolized by the participle, 'which is activated by touching', e. g. *érintőcsúszka* 'touch.PTCP-bar', is the virtual toll of a device that can be manipulated by touching. These structures therefore add further specification to the secondary participant of the process. As a consequence, the expressions' meaning shifts away from the prototypical construal of the process as a physical event (see Table 3, $p = 10$, $n = 6218$).

Table 3: Frequency data on the 'virtual secondary figure' category (C3) in HNC

	Number of tokens	**Proportion within category (%)**
érintőmező 'touch.PTCP-field'	4	40
érintőcsík 'touch.PTCP-stripe'	2	20
érintőcsúszka 'touch.PTCP-scroll bar'	2	20
érintőmenü 'touch.PTCP-menu'	1	10
érintőmutató 'touch.PTCP-pointer'	1	10

4.5 The profile is the result of the process

Falling even farther from the prototypical elaboration of the participle's process are semantic structures profiling abstract entities. These latter emerge as a result of the process of touching, e.g. an application that can be run and used by touching, as in the case of *érintőalkalmazás* 'touch.PTCP-application; touch.PTCP-app'. Hence semantic integration reaches a highly advanced level in the category. The processual meaning is elaborated by summary scanning (cf. Langacker 2008; 2009; 2016),[11] and

[11] In these cases, the process of touching is not elaborated as a specific physical activity; the process is a stable characteristic of the entity (which also has a processual character), hence it is construed schematically, it becomes "active and available as a simultaneously accessible whole for a certain span of processing time" (Langacker 2008: 83). This is the

the construction profiles the result of this process (see Table 4, p = 25, n = 6218).

Table 4: Frequency data on the 'result' category (C4) in HNC

	Number of tokens	**Proportion within category (%)**
érintőfókusz 'touch.PTCP-focus'	20	80
érintőfunkció 'touch.PTCP-function'	3	12
érintőalkalmazás 'touch.PTCP-app'	1	4
érintőszoftver 'touch.PTCP-software'	1	4

4.6 The profile is a complex process

Finally, those compounds are the farthest from the meaning of syntactic phrases which elaborate the participle's processual meaning by summary scanning, and profile a process or thing including the participle's process (TOUCHING) as a major conceptual component. For example *érintőszobrász* 'touch.PTCP-sculptor' profiles a sculptor who applies a holistic sculpting process including a salient sub-process of TOUCHING. Here, conceptual proximity is manifested as a part/whole relation, cf. *érintőstílus* ('touch.PTCP-style', 'a playing style on touch guitars which uses tapping'), or by way of metaphor, cf. *érintőgyémánt* ('touch.PTCP-diamond', metaphorical name for a touch-screen smartphone) (see Table 5, p = 22, n = 6218).

---

reason why we consider the semantic construal of PTCP's in this category to involve summary scanning. The temporal configuration of the processual meaning of Hungarian PTCP in syntactic structures as well as in compounds has not been investigated yet. According to our results, however, the following proposal seems to be reasonable: the more concrete is the process of the PTCP (by being a specific instantiation of an event type) in the whole meaning of the structure (e.g. in C1 or in C2 in the case of 'touching'), the more elaborated is the foregrounded process (indicated by the verb) through sequential scanning. This results in a more temporal meaning of the PTCP; the N component of the compound directs the conceptualizer's attention to one phase of the process. And, conversely, the schematization of the process indicated by the verb entails summary scanning and a less temporal meaning (without a clear figure–ground alignment of the processes in PTCP), which serves as vantage point for elaborating the meaning of the N component. Thus, a detailed examination of the processual character of PTCP in compounds can shed new light on the temporality of its meaning.

Table 5: Frequency data on the 'complex process' category (C5) in HNC

	Number of tokens	**Proportion within category (%)**
érintővezérlés 'touch.PTCP-control'	6	27.3
érintőtechnológia 'touch.PTCP-technology'	5	22.7
érintődesign 'touch.PTCP-design'	2	9.1
érintőstílus 'touch.PTCP-style'	2	9.1
érintőgyémánt 'touch.PTCP-diamond'	2	9.1
érintőszobrász 'touch.PTCP-sculptor'	2	9.1
érintőgörgetés 'touch.PTCP-scrolling'	1	4.5
érintőképesség 'touch.PTCP-capacity'	1	4.5
érintőgesztus 'touch.PTCP-gesture'	1	4.5

4.7 Comparing the frequency of the categories

The data show that the most frequent pattern in our sample is the composite structure profiling the secondary figure of the process (C2), whereas the lowest frequency is associated with expressions profiling a virtual secondary figure (C3, see Table 6).[12]

Table 6: The token and type frequency of the categories in HNC

Category	**Proportion – token frequency (%)**	**Proportion – type frequency (%)**
C1. profiling primary figure (instrument)	0.7	12
C2. profiling touched surface as secondary figure	98.3	46
C3. profiling virtual secondary figure	0.2	12
C4. profiling the result of the process	0.4	9
C5. profiling a complex process	0.4	21

[12] The importance of token frequency as a measure follows from our intuition, namely that the *érintő*-N compounds are relatively frequent expressions in contemporary Hungarian. The primary aim of the corpus study was to confirm this hypothesis, which is the reason why we supply token frequency data first.

The second category has the highest token frequency, which means that the members of this category underwent progressive entrenchment, and achieved unit status. Established units are those linguistic structures which become representative members of a category for a speech community. But from unitization it does not follow that the structure loses its analysability: the components may remain understandable in themselves for the conceptualizer. Consequently, a repeated and hence entrenched unit can become the baseline for new structures, as well as for further schematization (see Langacker 2008: 16–17). From the perspective of the latter cognitive process, type frequency is an important measure, specifying the number of different types of compounds representing each category in the sample. Type frequency can inform us about the variability of the category, and hence it tells us something about the productivity of the category: the higher the type frequency is, the higher the possibility to generate new tokens by this type (Barðdal 2008: 28; Bybee 2010: 67). On the other hand, a higher type frequency (the emergence of more and more types of a structure) facilitates the extraction of the commonalities inherent in multiple usage events, and therefore it can lead to schematization of the structure. As we can see from the data, not all *érintő*-N compounds are routinized individual expressions. The entrenchment of the constructions reaches the level of schematization: there are several constructional schemas in the background of compound formation, and while the stability of these schemas is not the same, they motivate new instantiations, consequently the process can be considered productive.

In this respect, the second category is again dominant. This means that the compound type profiling a touched surface as secondary participant is the most productive semantic schema for integration. While compounds profiling a virtual secondary figure have a low token frequency (the individual expressions are less entrenched as units), they produce almost as many compound types as the schema profiling the primary figure of the process. Exceptionally high is the type frequency of the schema profiling a complex process (compared to its token frequency). One reason for high frequency of types and low frequency of tokens in a category can be the emergence of a productive schema without a strong pattern of analogous compound formation; in this case, the process of forming new compounds is not item-based (analogy) but schema-based (elaboration), and we can assume an abstracted conceptual organization in the background. Additionally, the results show that such compounds easily undergo lexicalization. In other words, it seems that in the process of compound

formation we need to recognize a stage in which semantic change has no effect on the components' degree of integration (i.e. their unitizaton) but results in a decrease in their degrees of analysability and conceptual autonomy (as a function of the composite structure's lexicalization).

High token frequency indicates a high level of routinization in using specific patterns of compound formation. On the other hand, lexicalized compounds (the members of C5 in our case) have high type frequency (the second highest of the investigated pattern), but the types of the category are individual semantic structures. There is a relatively frequent pattern of types at work here with the abstract 'PROCESS – THING' meaning, instantiated by specific tokens. Lexicalization seems to be a very special case of compound formation regarding both productivity and frequency. High type frequency shows that there is a relatively productive schema in the background of the corpus data; however, it is a very abstract schema with the process of touching in its centre. We can assume that the schematization of more specific constructions may support the emergence of the highly abstract schema as well. On the other hand, lexicalized expressions do not have a prominent token frequency in our pattern; the reason for this can be that the unitization of these structures in the speech community is in its initial phase yet.

Which processes will be lexicalized with summary scanning by foregrounding the semantic component of TOUCH does not follow from the schema of TOUCHING. However, language users still find it convenient to employ complex, holistic names for the entities in question. In such cases, it may be presumed that TOUCHING has become an especially salient component of the thing denoted by the second part of the compound, which motivates the use of the compound construction. Compounding seems to be productive in these cases because the thing symbolized by the second component is accessed via the holistic conceptualization of TOUCHING (and not through the profiling of a component of the process). Therefore, such data may invite an analysis in terms of the control cycle as well (cf. Langacker 2009; 2016, see §6.2 for details). In conclusion, lexicalization cannot be derived schematically from semantic extensions of the verb *érint* 'touch'. However, compounding also has a conceptual motivation here, as it symbolizes a high degree of reconfiguration in component meanings. In other words, lexicalization suggests itself as a natural construal operation for naming certain entities, which explains the varied types within this category.

5 The data of the questionnaire

5.1 Analysability in the data gained by questionnaire

The analysis of corpus data explores the rich variability of *érintő*-N compounds. Our results clearly demonstrate the gradient, scalar nature of compounding patterns. A questionnaire study was designed to explore how PTCP-N structures are processed in general.

For the exploration of schemas, two expressions' data were relevant, those of *rimpuló tami* and *pivogolótami*. Recurring sound structures (*tami, rimpuló: rimpít, pivogoló: pivogolás*) generally prompted the informants to supply analyses. In the relationship between *rimpít* and *rimpuló*, meaning attribution usually incorporated the fact that *rimpít* can mean 'make rimp', whereas *rimpuló* can mean 'becoming rimp' (in Hungarian, verbs ending with *-ul* denote a change of state, those ending with *-ít* designate processes whereby some agent induces a change of state in another entity; *-ó* is the suffix of the *-Ó* participle).[13] In the case of each expression, a scale of analysability emerged, from complete lack of analysability (lexicalization) to in-depth analysability. We evaluated analysability on a scale of three degrees but with no intention to enforce clear-cut distinctions between the degrees. Degrees of analysability (based on the descriptions in meaning attributions, cf. examples of questionnaire data) are demonstrated below by responses to *pivogolótami* (see Table 7. The symbol I_n stands for the informant as a source of data.

[13] It must be added, however, that since the expressions are nonsensical, these analyses are only potentialities. The aim of the questionnaire was to explore the meaning initiating role of the *-ó* ending in a unipolar construction. In other words, we sought to find out whether the potential PTCP-N structure induced a bipolar (symbolic) interpretation of the nonsensical expression (whether the phonological structure evokes a potential semantic structure in the informants, see Langacker 2008: 16), and if so, in what proportion it was characteristic.

Table 7: Degrees of analysability for *pivogolótami*

Analysis	**Fully lexicalized meaning, lack of analysability**	**Lexicalization within the construct; partial analysis**	**Detailed analysis**
Example for meaning attribution	*ősi sumér tánc* 'ancient Sumerian dance' (I_6), *ez egy játék* 'this is a game' (I_9), *valamilyen tárgy vagy állat* 'some kind of object or animal' (I_{36})	*Pivogolótamból származik* '[something/ somebody that] comes from Pivigolótam' (I_{11})	*olyan tami, amely pivigol* 'a *tami* which *pivogol*s ('does *pivogol*ing') (I_8)
Grammatical form	N	N(pivogolótam)-i_{ADJ}	N(tami)$_{AG/NOM}$ V(pivogol)

5.2 A comparison of meaning attributions to *rimpuló tami* and *pivogolótami*

Both expressions showed a high degree of analysability, with detailed analysis reaching 88% in the case of *rimpuló tami* and 81% for *pivogolótami*. This suggests that on the basis of their linguistic forms, these expressions were considered to have a transparent structure. However, lack of analysis was more frequent for the one-word expression (17%) than for *rimpuló tami* (6%).

Table 8: Degrees of analysability for *rimpuló tami* and *pivogolótami*

Expression	***Rimpuló tami* PTCP N**	***Pivogolótami* N(PTCP-N)**
Detailed analysis	88%	81%
Partial analysis	6%	2%
Unanalyzed	6%	17%

With regard to the semantic relation between the process and the participant expressed by the noun, the descriptions reflect a clear difference between the two structures. The informants explicitly expressed (and in many cases highlighted as a motivating factor behind one-word spelling) the fact that with *pivogolótami*, the process (*pivogolás* 'pivogol.N_{DERIV}, the act of pivogoling') is not a one-off occurrence but rather a stable or at least regular feature of the thing referred to (e.g. 'a tool, a *tami*, that is inextricably linked to the fact that it *pivogol*s generally' [I_{45}]). Although this criterion also occurred once in the case of the phrase *rimpuló tami*, this is a much lower share compared to similar descriptions of the compound's meaning (see Table 9). The table only displays data where the informants explicitly

referred to a typical (constant, regular, frequent) or actual (occasional, ongoing) process (of *rimpulás* 'rimpuling' or *pivogolás* 'pivogoling') in their meaning attributions (n = 50).

Table 9: The process as a typical vs. occasional characteristic for N

	Rimpuló tami **PTCP N**		***Pivogolótami*** **N(PTCP-N)**	
The process (expressed by the PTCP)	Characteristic for N	Actual	Characteristic for N	Actual
	11 (22%)	16 (32%)	29 (58%)	1 (2%)

5.3 The constructions of *rimpuló tami* and *pivogolótami* on the basis of meaning attributions

As the expressions were nonsensical, the responses provide schematic constructional patterns that can motivate the elaboration of specific meanings. The constructions are highly similar for the two expressions. In what follows, we start off with the constructions of *pivogolótami,* and present constructional schemas and their proportions in meaning attributions by moving from active to increasingly passive structures (see Table 10).

For both expressions, semantic descriptions of the type 'a tami which rimpuls/pivogols; it is characteristic of the tami that it rimpuls/pivogols; the tami's function is rimpuling/pivogoling' are typical, i.e. the $N_{1AG\,/\,THEME}$ V_1 active construction. It is striking, however, that while the active AGENT-ACTION and the less active THEME-PROCESS relations predominate in interpretations of the phrasal pattern (93% in total), a much higher degree of variability is found in meaning attributions to the compound, with clearly active meanings having a more limited share (40%). In Table 10 the numbers in subscripts designate the processes and participants belonging together, e.g. the process V_1 has its primary participants as N_1. The semantic roles as well as the case forms are written also in subscripts after the nouns, e.g. $N_1(tami)_{AG/NOM}$ designates that the noun *tami* (expresses the primary participant of the process) has an agentive role and a nominative case.

Table 10: The distribution of active and passive construals

		Pivogolótami **N(PTCP-N)**		***Rimpuló tami*** **PTCP N**	
	Entity–process	Constructional patterns applied in meaning attribution	%	Constructional patterns applied in meaning attribution	%
active	agent and its action	$N_1(tami)_{AG/NOM}$ V_1(pivogol)	26	$N_1(tami)_{AG/NOM}$ V_1(rimpul)	64
		$N_1(tami)_{THEME/GEN}$ has the characteristic for $N_2(V_1(pivogol)\text{-ing})$	12	$N_1(tami)_{THEME/GEN}$ has the characteristic for $N_2(V_1(rimpul)\text{-ing})$	20
		-	-	$N_1(tami)_{THEME/NOM}$ is a V_1(rimpul)-ing type of N	9
	thing capable of acting	$N_1(tami)_{THEME/NOM}$ is able to do $N_2(V_1(pivogol)\text{-ing})$	2	-	-
	instrument of the action (used by the agent(s))	$N_1(tami)_{THEME/NOM}$ is for $N_2(V_2(pivogol)\text{-ing})$	18	-	-
		with $N_1(tami)_{INSTR/INS}$ they V_2(pivogol); with $N_1(tami)_{INSTR/INS}$ it is possible to V_2(pivogol)	10	-	-
passive		$N_1(tami)_{INSTR/ACC}$ they use for $N_2(V_2(pivogol)\text{-ing})$	6	-	-

The different degrees to which phrases and compounds are associated with active meanings explain why the figures profiled by the nouns in phrases vs. compounds were considered as persons or physical objects (more specifically tools) with different degrees of likelihood. In meaning attributions to *pivogolótami*, the entity was described as a physical object in 42% of cases, whereas *rimpuló tami* had a corresponding score of only 8%.

5.4 Schemas associated with the N(PTCP-N) pattern and motivating factors behind compounding on the basis of the questionnaire study

In the case of one-word spelling, informants were more likely to refrain from analysis (see criterion 1 in Table 11), and more often chose to provide descriptions of lexicalized meanings. One-word spelling activated the interpretative strategy that perceived components of the expression could be processed at a higher degree of conceptual integration[14] in comparison to two-word spelling. In accordance with this iconic motivating principle, relations marked by the adjacency of PTCP and N were interpreted as tighter, more stable and more integrated with one-word than with two-word spelling, where the distance is greater between the two component structures (see criterion 2 in Table 11).

In close correlation with the motivating factor of higher conceptual integration, the construction types informing meaning attributions were more varied for compounds than for phrases (see criterion 3 in Table 11).

The PTCP-N structure prototypically receives an active interpretation. However, an active reading of THING–PROCESS relations is more dominant with two-word than with one-word spelling (see criteria 4 and 5 in Table 11). Thus, an increase in integration (semantic lexicalization) makes it less likely that the THING receives an active interpretation. In the PTCP-N structure, the N typically denotes a PHYSICAL OBJECT, the primary figure of the PROCESS (ACTION), i.e. its AGENT or INSTRUMENT.

Table 11: Summary of results gained by the questionnaire study

	Rimpuló tami **PTCP N**	***Pivogolótami*** **N(PTCP-N)**
1. Unanalyzed	6%	17%
2. PTCP is characteristic for N	11%	58%
3. Number of types of constructional patterns	3	6
4. N is construed as an active participant	84%	40%
5. N is construed as a passive participant	0%	6%

In the N(PTCP-N) structure, N is typically an active participant, the primary or secondary figure of the process, and it corresponds to the subject or means adverbial of the verbal stem of the participle. In our data, it corresponded to the object in only 6% of cases. We found no data in which the thing denoted

[14] See also as conceptual coherence or cohesion (cf. Barðdal 2008).

by the noun would bear a locative or other circumstantial relation to the verbal stem.

When it comes to the compound *pivogolótami*, only two categories were supported by the data with regard to profiling. The expression profiles the primary figure ('the thing or person that pivogols') in fully active patterns (40%), and the secondary figure in the role of INSTRUMENT in transitional (partially active) (18%) and passive constructions (16%) (see Table 10, 10% + 6%). The informants' interpretations suggest that the compound does not profile the secondary figure either as PATIENT (the thing that undergoes the action of pivogoling) or in any other thematic role. We will return to these observations when they are compared with the results of our analysis of the corpus data (see §6.1).

6 The general semantic model of PTCP-N structures

6.1 Schematic constructions behind the corpus data

The analysis of corpus data explores the schematic structures that arise in elaborations of a single processual meaning (that of TOUCHING) with regard to the profiling of various participants. By contrast, the investigation of questionnaire data reveals constructions that license linguistic patterns as recognized by language users.

The starting point for exploring correlations[15] is that the PTCP-N construction has a basically active meaning. This has been confirmed by meaning attributions in the questionnaire, and presumably our data would converge even more to this pattern if we studied other participial components of compounds (e.g. *futó* 'run.PTCP, running' or *sikló* 'glide.PTCP, gliding'; however, here again there would be departures from active construal, e.g. by profiling the PLACE as secondary figure).

The ubiquity in the corpus of participial components in passive compounds suggests that there is a schematic construction in the background which deviates from the usual (phrasal) structure. Put differently, *érintőképernyő* 'touch.PTCP-screen' and similar expressions seem to instantiate a construction that is distinct from the baseline PTCP-N schema. Our results of the corpus study further imply that there is more than a single construction at work: compounds with *érintő-* may be motivated by several

[15] By the term correlation we do not mean statistical connection here; instead, we would like to refer to the parallelism with which the two phenomena (pattern of compounds in the corpus and constructions gained by the questionnaire) can be related to each other.

schematic structures specifying the relationship between the process and its participants in different ways.

On comparing corpus data with construction-related data, we assumed that the frequency of constructions derived from meaning attributions to nonsensical expressions cannot be correlated directly with the token frequencies of particular categories for corpus data with *érintő-* 'touching'. This is because the general PTCP-N schema is situated at the active pole of construal, whereas the linguistic data under study give evidence of the emergence of specific passive meanings. Therefore the two frequency distributions should not match, with the constructions extracted from questionnaire data showing a different manifestation of passivization than compounds with *érintő-* as their initial components. At the same time, we also expected that both distribution patterns would highlight the variability of constructions, and moreover, that the frequency of passive constructions (derived by questionnaire) would correlate with the type frequency of passive categories in the corpus data.

Thus, one criterion of comparison (and of the identification of distinct schemas) is the passive vs. active character of constructions. Constructions of the nonsensical expression *pivogolótami* are arranged on the active/passive scale into the distribution shown earlier (Table 10). To recapitulate our findings, active constructions account for 40% of the full sample, and are internally varied, with the profiled entity typically accomplishing the process (38%) or having the capacity to accomplish it (2%). Constructions classifiable as passive display similar internal variability, but their overall frequency in meaning attributions is lower (34%), of which unequivocally passive structures have a share of 6%.

The semantic categories of the corpus data on the basis of their degrees of semantic integration show a different distribution with regard to the active/passive continuum (see Table 12).

Table 12: The semantic categories of the corpus data in the active/passive continuum

	Category	**Token frequency (%)**	**Type frequency (%)**
active	C0: the profile is the primary figure as AG	0	0
partially active	C1: the profile is the primary figure as INSTR	0.7	12
passive	C2: the profile is the secondary figure as touched surface	98.3	46
	C3: the profile is a virtual secondary figure	0.2	12
	C4: the profile is the result of the process	0.4	9
	C5: the profile is an entity including the process	0.4	21

The proportion of passive semantic integrational schemas stands out, thus the corpus data match the variability of constructional schemas that we found in meaning attributions. In the corpus data, the sample is much more heterogeneous in the passive domain of the scale, with four out of five categories showing passive, or at least partially passive meaning.

Hence, our hypothesis about the scalarity of active–passive construal proved to be correct: passive meaning as a semantic motivating factor is not a homogeneous phenomenon; rather, the passive construal of a process has varied manifestations.

It follows from the discrepancy between data types that the categories of corpus data and the constructions established by the questionnaire cannot be directly compared on a schema-by-schema basis. Whereas meaning attributions rely on components to circumscribe the meaning of each expression, corpus data do not supply comparable results; a given category, e.g. the one represented by *érintőceruza* ‘touch.PTCP-pencil’, may correspond to a variety of constructional schemas in responses to the questionnaire (see Table 13).

Table 13: Documented constructions of *pivogolótami* and possible interpretative/explanatory constructions for *érintőceruza* 'touch-PTCP-pencil'

Entity–process relationship, scale of activity		**Constructions of *pivogolótami***	**(%)**	**Possible constructions of *érintőceruza***
active	agent and its action	N_1(tami)$_{AG/NOM}$ V_1(pivogol)	26	N_1(pencil)$_{?AG/INSTR/NOM}$ that V_1(touches) N_2
		N_1(tami)$_{THEME/GEN}$ has the characteristic for N_2(V_1(pivogol)-ing)	12	N_1(pencil)$_{THEME/GEN}$ has the characteristic of touching
	thing capable of acting	N_1(tami)$_{THEME}$ is able to do N_2(V_1(pivogol)-ing)	2	N_1(pencil))$_{THEME}$ is able to touch, to do the touching [process]
	instrument of the action (used by the agent(s))	N_1(tami)$_{THEME}$ is for N_2(V_2(pivogol)-ing)	18	N_1(pencil))$_{THEME}$ is for touching
		with N_1(tami)$_{INSTR}$ they V_2(pivogol); with N_1(tami)$_{INSTR}$ it is possible to V_2(pivogol)	10	with N_1(pencil)$_{INSTR}$ they touch N_2; with N_1(pencil)$_{INSTR}$ it is possible to touch N_2
passive		N_1(tami)$_{INSTR/ACC}$ they use for N_2(V_2(pivogol)-ing)	6	N_1(pencil)$_{INSTR/ACC}$ they use for touching

Thus, when the schematic constructions gained by the analysis of questionnaire data are brought into correspondence with corpus categories, syntactic construal (see e.g. *olyan tami, ami pivogol* 'a tami which pivogols' (I_2), *pivogolást végző tami* 'tami doing pivogoling' (I_3) and *egy tami, ami* […] *általában pivogol* 'a tami which generally pivogols' (I_6)) and the thematic roles associated with the event (AG, THEME, INSTR) are both important criteria. However, as a consequence of the greater schematicity of constructions, a given construction type may correspond to several categories of corpus analysis (Table 14). This gives the study its bidirectionality and dynamicity; not only do we look at instantiations from the perspective of schemas, but also feedback the lessons of token analysis to the level of schematic constructions. The fact that our corpus data cannot be reduced to the construction types established by questionnaire responses suggests that novel schemas emerge in language use, or else linguistic expressions are motivated by the extension of existing schemas, i.e. conventionalization is under way.

On a micro-level, a pre-requisite for comparing the two datasets is to provide questionnaire data with a semantic analysis in terms of profiling (see Table 14).

Table 14: A comparison of constructional schemas with regard to profiling

	Corpus data (HNC)		**Questionnaire**	
Category	Token frequency (%)	Type frequency (%)	Construction	%
C0: the profile is the primary figure as AG or THEME	0	0	$N_1(tami)_{AG/NOM}$ that pivogols; $N_1(tami)_{THEME/GEN}$ has the characteristic of pivogoling; $N_1(tami)_{THEME/NOM}$ is able to do pivogoling	40
C1: the profile is the primary figure as INSTR	0.7	14	$N_1(tami)_{THEME/NOM}$ is for pivogoling; with $N_1(tami)_{INSTR/INS}$ they pivogol; with $N_1(tami)_{INSTR/INS}$ it is possible to pivogol; $N_1(tami)_{INSTR/ACC}$ they use for pivogoling)	34
C2: the profile is the secondary figure as touched surface	98.3	44	-	-
C3: the profile is a virtual secondary figure	0.2	12	-	-
C4: the profile is the result of the process	0.4	9	-	-
C5: the profile is a complex process	0.4	21	-	-

This analysis further modifies the picture when it comes to the active/passive dimension in the meaning of compounds. In particular, it can be established that in the corpus data even the category implementing active meaning is only partially active (as it does not profile the agent). By contrast, in the questionnaire data, even the fully passive construction is instantiated in such a way that it profiles the primary figure of the process. This supplies a further argument for interpreting the construal of passive meaning as a key motivating factor for the emergence of compounds. Thus, an emergent schema can be posited as licensing structure behind the relevant compounds in our corpus data.

In conclusion, constructional analysis may aid the classification of actual occurrences; however, the established constructions cannot be directly matched with the pattern emerging from corpus data. A construction considered as frequent need not be frequently instantiated in actual language use, and a construction that informants regard as rare may produce a variety of compound types in the corpus.

The semantic categories supported by corpus data confirm the existence and variability of the active/passive continuum. The semantic schemas may be classified in terms of the specificity (degree of elaboration) of processual meaning, in basically the same way as shown by the active/passive scale. From this it follows that construing a passive meaning (in varied ways) clearly motivates the emergence of compounds. At the same time, compounding as an operation involves increasingly tight and specific patterns of semantic integration, leading to the emergence of new constructions as well. In other words, an increase in conceptual proximity between the components may manifest itself in semantic integration, and also in novel constructional schemas.

6.2 The meaning of PTCP-N as control cycle

Our complex analysis of the variability of *érintő*-N compounds in the corpus has generalizable results. In this section, we present a comprehensive account of PTCP-N compounds by building on the findings, making crucial use of Langacker's concept of the control cycle.

In Cognitive Grammar, semantic extension is assumed to occur in four phases (cf. Langacker 2009: 130–135; 2016), with an agent gradually extending his/her control to an entity. The starting point is the baseline of the structure, where an agent controls certain entities (located in its immediate scope). This is followed by the potential phase, in which control over a new entity arises as a possibility. Through reinforcement by repetition, this possibility becomes a stable property of the agent, which marks the action phase of the process. Finally, in the result phase, the newly controlled entity is incorporated into the agent's immediate scope. Figure 2 below represents the process of compound formation in a schematic way. Small empty circles above designate the entities being under the actor's control. They are involved in schematic (prototypically force-dynamic) actions, thus they do not represent actual objects in a specific situation. However, from the potential phase the type of the controlling action becomes specified, namely as the act of touching something.

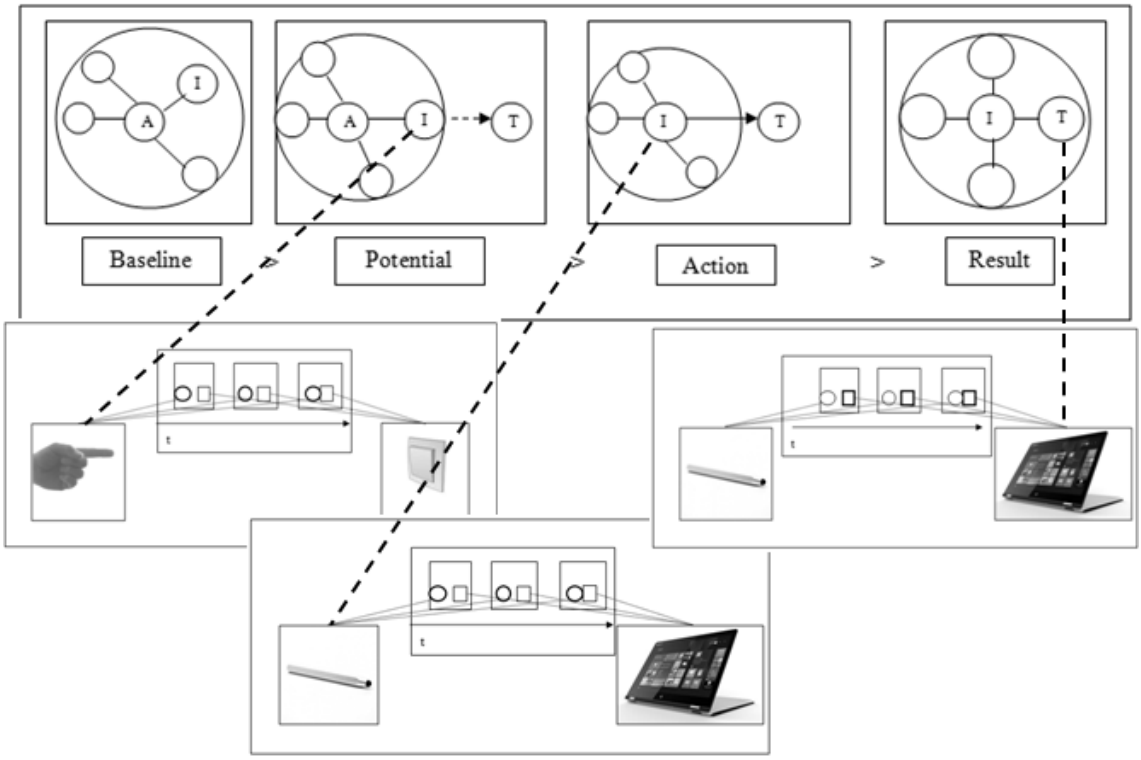

Figure 2: The development of *érintő*-N as a control cycle

This model is well-suited to the explanation of meaning extension based on the process of touching. The baseline corresponds to the situation in which the agent controls a tool which can accomplish certain actions. In the potential phase, the tool is involved in a single, transient touching event (as expressed in the phrasal construction by the PTCP modifier). However, when it becomes motivated for this possibility to be repeatedly seized upon (e.g. because the tool is generally suitable for accomplishing the process profiled by the participle), the process develops into a recurring occurrence, and the tool is endowed with the property of accomplishing it. Finally, the tool replaces the agent, and through the process of touching it is able to control other entities (see Figure 2).

In the picture, the potential, action and result phases of the cycle are illustrated by three different situations. When a tool is suitable for manipulating an entity, the cycle is set in motion, but it only comes to completion when this kind of manipulation becomes a salient property (even a raison-d-être) for the tool. In the absence of this, phrasal patterns become conventional on the phonological pole. If, however, the tool is involved in repeated acts of manipulation, and this even becomes its main purpose, the cycle proceeds to the action phase. Here, the agent's place is assumed by the

tool (via metonymization), and the latter is also profiled by the construction, allowing for the emergence and entrenchment of compounds. Compounds in the first category of our corpus data are characteristic instantiations of this phase of the cycle. However, spelling varies when the tool's regular association with the process is not yet widely known in the community, and the compound is not yet sufficiently stable as a unit in the particular situation. After this, the cycle reaches its result phase. On the one hand, this is indicated by further shifts of attention (from the manipulating tool to the manipulated entity, as shown by the second and third categories of our corpus data). Once the cycle has finished, the emergent meaning becomes a productive semantic schema with high type frequency (and with high token frequency of the constructions, but not necessarily of the individual expressions) as a consequence. On the other hand, the process of manipulation gets schematized, with less and less concrete, more metaphorical interpretations of control arising as illustrated by examples in the fourth and fifth categories of our corpus data. The course of the cycle results in the emergence of more or less specific constructional schemas (regarding the elaboration of the process TOUCHING). But the emergence of these schemas also supports the continuation of the cycle itself. Consequently, the process of the cycle and the entrenchment of constructional schemas reinforce each other mutually.

The end phase of the cycle establishes a highly schematic semantic structure (TOUCHING – TOUCH-RELATED ENTITY in this case), which, however, does not result in entrenched units, and shows high type frequency with a limited number of tokens. Thus, the last phase of the cycle triggers a process of lexicalization, which is beyond the cycle. It is worth noting at this point that the model based on the control cycle is sufficiently general to account for the emergence of further PTCP-N patterns.[16] Of course, there are varied ways of establishing control, and the controlling tool and the controlled entity also vary with the nature of the process profiled by the

[16] Our argumentation relying on the control cycle helps us to refine a previous claim of Bundgaard et al. (2006: 387), namely that "*X* an *Y* may combine aspects of a teleological process freely." By adopting the cycle as a motivating factor behind compound formation and the emergence of constructional schemas, we propose that the actual pattern of compounds in a corpus is not a free variation of conceptualization (being governed only by creativity and the context), but the result of consecutive changes in conceptual organization. Although the members of the investigated compound category (*érintő*-N) occur simultaneously in the corpus, they correspond to particular phases of a cognitive and linguistic reorganization process.

participle. For example, in the case of physical movement, the tool first assumes the position of the controlled entity, as in *futócipő* 'run.PTCP-shoe, running shoe', then the local participant of the movement, cf. *futópad* 'treadmill', lit. 'run.PTCP bench, running bench', and finally the manipulation offered by movement is only schematically part of the construction's meaning, as shown by *futótárs* 'run-PTCP-partner, running partner'. It goes without saying that changes in the motivating frames and shifts between them also require characterization specifically for each process as elaborated in discourse.

Nevertheless, in the case of compounds with an initial PTCP component, the cycle is always implemented (whichever participant is profiled). Thus, this model interprets the emergence of compounds as a process in which semantic integration between the components occurs as part of a more abstract process of schematization. The variability of corpus data is motivated by different phases of this schematization process. Hence, although the categories can be studied by themselves, they cannot be regarded as independent from the other categories and from the overall process of semantic extension. The questionnaire data, for their part, indicate that in the case of nonsensical expressions, the cycle only begins but does not necessarily reach its endpoint, which is why active construal has such a high share in the responses. It is also clear, however, that separate constructions emerge in language use in parallel with the default sequence of the cycle.

The model of control cycle has two advantages compared to the general event frame (or to the purpose-oriented action theory, see Bundgaard et al. 2006: 385–388). One is its specificity, the other is its dynamic character. On the one hand, the phases of the cycle can model the aspects of developing control over an entity (as a schematic purposeful action), with the benefit of explaining the entrenchment of particular expressions as established linguistic units of one or another phase of the cycle. On the other, the control cycle is as general as the event frame model; however, it offers a processual explanation for compound formation, hence it can also explain the emergence and availability (and even the relative productivity) of intermediate constructional schemas.

7 Conclusions

In this paper, we presented a new approach to compounds, arguing that compositional descriptions do not efficiently account for their emergence. We interpreted the emergence of PTCP-N compounds from phrases as guided

by a complex schematization process which makes varied instantiations possible and results in the emergence of schematic constructions.

By subjecting our corpus data to both quantitative and qualitative analysis (relying on Cognitive Grammar), we described the rich variability of *érintő*-N compounds. In addition, a questionnaire study was used to confirm the emergence of schematic constructions and the hypothesis of passive construal being a key motivating factor behind compounding. Employing the control cycle model of Cognitive Grammar, we finally proposed a comprehensive account for the emergence of compounds with a PTCP initial component.

Although we did not implement a compositional description, this is not to deny that usage events involve a series of operations for integrating semantic components. Compounds were presented as semantic constructions rather than as syntactic structures; however, by starting off with phrases, we also allowed room for syntactic analysis. Our results clearly demonstrate the gradient, scalar nature of compounding patterns. Scalarity was established not by a study of data involving different components, nor did we carry out psycholinguistic experiments. Rather, our argument rests on the analysis of a large amount of corpus data on compounds with the same initial component.

The control cycle model interprets the emergence of compounds as a process in which semantic integration between the components occurs as part of a more abstract process of schematization. With accepting the cycle as motivating factor of compound formation and the emergence of constructional schemas we propose that the actual pattern of compounds in a corpus is not a free variation of conceptualization (being governed only by creativity and the context), but the result of a consecutive change in conceptual organization. Though the members of the investigated compound category (*érintő*-N) occur simultaneously in the corpus, they can be arranged along with a cognitive and linguistic reorganization process.

In the future, it will be worth extending the analysis to other compounds with a PTCP initial component, which may bring refinements to the comprehensive model. Beyond this, it will be necessary to use experimental measurements with a view to identifying the roles of passive construal, degrees of elaboration in the process profiled by the participle, and the salience of particular components.

Acknowledgements

This paper was supported by the Hungarian National Research, Development and Innovation Office, under project number 129040, "Verbal constructions in Hungarian. Research in usage-based construction grammar", led by Gábor Tolcsvai Nagy.

References

Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. (Constructional approaches to language 8). Amsterdam: John Benjamins.

Benczes, Réka. 2006. *Creative compounding in English: The semantics of metaphorical and metonymical noun-noun combinations*. Amsterdam: John Benjamins.

Bundgaard, Peer F. & Ostregaard, Svend & Stjernfelt, Frederik. 2006. Waterproof fire stations? Conceptual schemata and cognitive operations involved in compound constructions. *Semiotica* 161. 363–393.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Dirven, René & Verspoor, Marjolijn. 2004. *Cognitive exploration of language and linguistics*. Amsterdam: John Benjamins.

Heyvaert, Liesbet. 2012. Compounding in cognitive linguistics. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of cognitive linguistics*, 233–254. Oxford: Oxford University Press.

Kiefer, Ferenc. 2000. A szóösszetétel [Compounds]. In Kiefer, Ferenc (ed.), *Strukturális magyar nyelvtan* 3 [Hungarian structural grammar 3]. 519–567. Budapest: Akadémiai Kiadó.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

—— 2009. *Investigations in cognitive grammar*. Berlin: Mouton de Gruyter.

—— 2016. Baseline and elaboration. *Cognitive Linguistics* 27(3). 405–439.

Oravecz, Csaba & Váradi, Tamás & Sass, Bálint. 2014. The Hungarian gigaword corpus. *Proceedings of LREC*, 1719–1723. (http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf) (Accessed 2018-03-16.)

Ryder, Mary Ellen. 1994. *Ordered chaos: The interpretation of English noun-noun compounds*. Berkeley: University of California Press.

Scalise, Sergio & Fábregas, Antonio & Forza, Francesca. 2009. Exocentricity in compounding. *Gengo Kenkyu* 135. 49–84.

Ungerer, Friedrich. 2007. Word-formation. In Geeraerts, Dirk & Cuyckens, Hubert (eds.), *The Oxford handbook of cognitive linguistics*, 650–675. Oxford: Oxford University Press.

Contact information:

Nóra Kugler
Gábor Simon
ELTE Faculty of Humanities
H-1088 Budapest, Múzeum krt. 4/A
Hungary
e-mail: kugler.nora(at)btk(dot)elte(dot)hu
simongabor(at)caesar(dot)elte(dot)hu