Infinitive-gerund alternations in Polish post-verbal complementation: A corpus-based study

AGNIESZKA KALETA

Abstract

The paper constitutes a contribution to semantically based research on the so called 'syntactic alternations', defined as "structurally and/or lexically different ways to say functionally very similar things" (Gries 2017: 8). More specifically, the study is concerned with Polish verbal predicates which accept both the gerundive and infinitival complement, with the difference between these two alternatives being a matter of construal or focus rather than any clear-cut semantic contrasts. Taking as its point of departure the cognitive linguistic thesis that a change in form always entails a change in meaning, the present paper seeks to determine the factors that significantly affect the choice of one or the other complement type in situations where both are fully acceptable. The study provides an in-depth, corpus-based analysis of three selected pairs of constructions, and offers a new theory (referred to as *reduced-increased involvement theory*) to explain the choices that Polish speakers make with respect to the two constructions in question.

Alternacje gerundium-bezokolicznik w polskojęzycznych konstrukcjach dopełnieniowych: Podejście korpusowe

Abstrakt

Niniejszy artykuł wpisuje się w nurt badań nad semantyką tzw. 'alternacji składniowych', które można zdefiniować jako alternatywne sposoby wyrażania podobnych treści przy pomocy odmiennych środków składniowo-leksykalnych (Gries 2017: 8). Przeprowadzone badanie dotyczy polskojęzycznych czasowników, które współwystępują z dwoma różnymi formami dopełnieniowymi – gerundium (rzeczownikiem odczasownikowym) i bezokolicznikiem, tworząc tzw. pary minimalne, a więc konstrukcje odnoszące się do tej samej sytuacji, lecz różniące się pod względem sposobu jej konstruowania. Przyjmując za punkt wyjścia kognitywną tezę, że zmiana formy zawsze oznacza zmianę znaczenie, artykuł podejmuje próbę określenia czynników semantycznych wpływających istotnie na wybór dopełnienia gerundialnego lub bezokolicznikowego w sytuacjach gdy obydwie formy są równie akceptowalne. Artykuł przedstawia szczegółową analizę trzech wybranych par konstrukcyjnych przeprowadzoną w oparciu o metodologię korpusową, oraz proponuje nową teorię wyjaśniającą wybór formy dopełnieniowej przez natywnych użytkowników języka polskiego (nazwaną wstępnie „teorią ograniczonego-wzmożonego zaangażowania").

1. Introduction

Research on complement clauses has been thriving since the 1970s. While the early generative studies were preoccupied

mainly with the syntactic aspects of complementation, the more recent functional approaches have concentrated primarily on issues of semantics, offering many insightful observations about the relationship between the form and function of different types of complement clauses (cf. Bresnan 1979; Givón 1980, 2001; Noonan 1985; Wierzbicka 1988). The issue of syntax-semantics interface is also an important theme in complementation research inspired by the framework of construction grammar: a family of different, yet related theories, which define grammar as a vast, monostratal repository of *constructions*, i.e. learned pairings of form with meaning (cf. Goldberg 1995, 2006; Langacker 1987, 1991, 2008). Hence, when seen from the constructional perspective, different types of complement clauses represent constructions in their own right, which entails that their distribution is not random or arbitrary but semantically motivated. One way of approaching this matter by constructional grammarians has been by surveying all of the verbs that take a certain kind of complement clause and establishing whether the verbs in each set share any semantic features which might represent a unitary value of the different complement types (cf. Dirven 1989; Smith 2008; Taylor and Dirven 1991). Another research strategy has been to focus on the verbs that take two or more complement types, usually without any obvious differences in meaning. This type of constructions can be subsumed under the general heading of syntactic alternations, a topic of much interest to construction grammarians over the last decade or two (cf. Gries 2017; Gries and Stefanowitsch 2004; Kaleta 2014; Rudanko 2021; Stefanowitsch and Gries 2003). The present study constitutes a contribution to this latter strand of research by focusing on alternations between the gerund and the infinitive, which can be found in post-verbal complement clauses in present day Polish. While there is ample literature on infinitive-gerund alternations in English complementation (cf. Deshores 2015; Deshores and Gries 2016; Kaleta 2012; Rudanko 2017, 2021; Smith and Escobedo 2001), the corresponding research on data from other languages is very scarce or even

non-existent, the result of which is that our hitherto knowledge on this type of alternations is heavily biased by the patterns found in the English language. The present paper aims to fill this gap by providing an analysis of complement patterns of selected Polish verbs that accept both the infinitival and the gerundive complement, as illustrated with the examples below:[1]

(1a) Polacy chcą wyjść z UE.
     Lit. 'Poles want to leave EU.'
(1b) Polacy chcą wyjścia z UE
     Lit. 'Poles want leaving EU.'
(2a) Zaczął pisać powieści w wieku 10 lat.
     Lit. 'He started to write novels at the age of 10.'
(2b) Zaczął pisanie powieści w wieku 10 lat.
     Lit. 'He started writing novels at the age of 10.'
(3a) Zakazali nam chodzenia do kościoła.
     Lit. 'They forbade us going to church.'
(3b) Zakazali nam chodzić do kościoła.
     Lit. 'They forbade us to go to church.'

While it is true to say that in certain cases (such as the ones above) the two complement patterns can be used more or less interchangeably, there are numerous instances in which the substitution of one form for another produces clear-cut differences in meaning. Consider (4a) and (4b):

---

[1] I use the term 'gerund' to refer to nominalizations also known as *substantiva verbalia* – in Polish they are derived with a fully productive -*nie*/-*cie* suffix (e.g. *pisanie, mycie*). *Substantiva verbalia* should be distinguished from *substantiva deverbalia* – while the former are morphologically regular and can be formed from nearly all verbs, the latter are much more idiosyncratic and can be derived only from a very limited number of verbs (cf. Krzeszowki 1990: 208). Krzeszowki describes *substantiva verbalia* (e.g. *pisanie, mówienie*) as being 'more nouny' than *deverbalia* (e.g., *pisarstwo, mowa*) and compares them to Lees' Gerundive Nominals.

(4a) Polacy chcą oddzielenia edukacji od religii.
  Lit. 'Poles want separating education from religion.'

(4b) Polacy chcą oddzielić edukację od religii.
  Lit. 'Poles want to separate education from religion.'

Also, there are cases where the substitution of one form with another produces clearly unacceptable expressions, as illustrated in (5b):

(5a) Zaczęło się odliczanie do Świąt Bożego Narodzenia.
  Lit. 'The counting down to Christmas has begun.'

(5b) *Zaczęło się odliczać do Świat Bożego Narodzenia.
  Lit. 'To count down to Christmas has begun.'

The existence of such restrictions clearly indicates that the two complement patterns under consideration are semantically distinct, i.e. each of them makes its own semantic contribution to the meaning of the whole utterance. Hence, the main question that guides the present research is the following: what is the semantic import of the infinitival and gerundive complement? Differently put – what are the factors that motivate the choice of one complement type over the other in cases where both are accepted by a given matrix predicate? This question is addressed with corpus-based methods, which are used to gain an insight into the distributional patterns of the two complement constructions in question. More specifically, the paper presents three case studies that explore in some detail the contextual linguistic factors that influence the choice of one or the other complement type by three semantically distinct verbs: a verb of volition – *chcieć* (want), a verb of aspect – *zaczynać* (begin), and a negative directive – *zakazać* (forbid) (cf. examples 1-3). Underlying this research is Harris's (1954) *distributional hypothesis*, which presumes the existence of correlations between distributional characteristics of linguistic structures and their functions/meanings.

The paper is structured as follows. Section (2) outlines the details of the methodological procedure followed in this study. Sections (3)-(5) present the results of the three case studies, i.e. they determine and discuss the distributional features that significantly distinguish the infinitival constructions from the corresponding gerundive constructions. Section (6), on the other hand, provides a qualitative interpretation of the distributional/ frequency data presented in the previous sections and offers a theory called *reduced-increased involvement theory* to explain the contrasts between the two constructions under investigation. More specifically, it is argued that the infinitive-gerund alternations encode different degrees of the matrix clause agent's involvement in the event described in the complement clause. It is shown that the gerundive constructions present the matrix agent(s) as being 'less involved' or 'less salient' and hence less in control over the complement event than the agents of the infinitival construction, who exert more control or are more directly affected by the action described in the complement clause. The last section concludes the discussion and offers some prospects for future research.

2. Methodological procedure

As pointed out in the introductory section, the study is situated within the framework of functional-cognitive grammar, which posits that all units of language, including syntactic forms such as the infinitive or gerund, are constructions in their own right, i.e. distinct form-meaning pairings, which means that their distribution is semantically motivated. This also entails that there are no fully synonymous constructions as "a difference in form always spells a difference in meaning" (Bolinger 1968: 127). Taking this assumption as a point of departure, the present study uses corpus data and corpus-based methods to argue for the semantic distinctiveness of the two nearly-synonymous complement constructions, as introduced in section (1): [verb + the infinitive] and [verb + the gerund]. The methodological

procedure consists of two steps. First, all the verbs taking the infinitival complement and the verbs complemented by the gerund have been extracted from the corpus. The corpus used for this study is the Polish Web 2012 corpus, available via Sketch Engine software.[2] Since this is a very large database consisting of circ. 812,818,518 words, a smaller sub-corpus of 43,226,158 was created in order to make the data more manageable and available for hand-editing if required. All the relevant patterns were extracted using the CQL (corpus query language) function of the Sketch Engine. Table 1 presents the codes used to extract the constructions in question, together with the corresponding token and type frequencies.

Table 1

Codes used for data extraction from Polish Web 2012

construction	CQL code	Tokens	Types
Verb+infinitive	[tag="V.*"][]{0,2}[tag="inf.*"]	463312	105
Verb+gerund	[tag="V.*"][]{0,2}[tag="ger.*"]	16712	191

In the next step, the verbs appearing with both the infinitival and gerundive complement were determined using a CQL code. Table 2 presents the verbs that occur more often with the gerund than the infinitive, while Table 3 lists the predicates that favour the infinitival complement over the gerundive one. The tables give the raw frequencies of each of the two constructions, as they occur with a given matrix verb.

---

[2] https://www.sketchengine.eu/

Table 2

Verbs with the higher incidence of gerundive complements

lemma	Inf.	Ger.
umożliwić/umożliwiać (enable)	5	762
rozpocząć/rozpoczynać (begin)	55	379
zaprzestać/zaprzestawać (stop/cease)	48	214
zakazywać/zakazać (prohibit)	65	191
zalecać/zalecić (advise/suggest)	23	104
zlecić/zlecać (order/instruct)	2	61
sugerować (suggest)	26	74
życzyć (wish)	39	40

Table 3

Lemmas with the higher incidence of infinitival complements

lemma	Inf.	Ger.
chcieć (want)	51580	190
zacząć/zaczynać (begin)	29982	123
próbować/spróbować (try)	11723	29
woleć (prefer)	2884	21
lubić (like)	5015	69
pomóc/pomagać (help)	4140	94
pragnąć (desire)	2383	47
uczyć/nauczyć (teach)	2778	140
uwielbiać (love)	1055	21
planować/zaplanować (plan)	820	214
proponować/zaproponować (propose)	750	301
nakazać/nakazywać (order/require)	508	139
polecić/polecać (tell/order)	452	148
skończyć/kończyć (finish)	370	120
obiecać/obiecywać (promise)	359	58
zabronić/zabraniać (forbid)	357	216
kochać (kochać)	221	21

Given the space limitations of the present paper, only three predicates have been selected for a detailed consideration, each representing a different semantic type: *chcieć*, *zaczynać* and *zakazać*. The analytical procedure at this stage also consisted of two steps. First, a sample concordance of each verb has been extracted from the sub-corpus and subjected to a preliminary qualitative inspection. Next, the concordance lines have been coded for the features that could potentially be significant in distinguishing the gerundive and the infinitival construction. These data were then subjected to the chi-square test in order to determine whether there is an association between a given variable and the constructional choices. Once the significant associations have been determined, the frequency data have been subjected to qualitative interpretation, on the basis of which a hypothesis concerning the semantic import of the constructions in question has been formulated. As pointed out in the introductory section, this approach is based on the distributional hypothesis, according to which the linguistic context in which a word occurs determines its meaning. Given that cognitive-functional models of grammar regard all units of language, irrespective of their degree of morpho-syntactic complexity, as form-meaning pairing, the distributional hypothesis can be easily extended to cover syntactic units. Hence, the assumption underlying this research is that the distributional properties of a construction are indicative of its semantic properties, which is to say that linguistic/contextual features provide the 'clues' to constructional semantics.

3. Case study 1: *Chcieć* and its complements

Even a cursory look at the relevant concordance lines makes it clear that the complement choices made by this basic verb of volition are significantly influenced by whether the subjects of the main clause and complement clause are coreferent or non-coferent (i.e. whether they refer to the same individuals or to different individuals). As it appears, in the former case the

infinitive is selected, whereas in the latter case the gerundive complement is the preferred choice. Consider the following example:

(6) Chcemy całkowitego *oddzielenia* edukacji od religii.
   Lit. 'We want the complete separation of education from religion.'

In this example the ones who want the separation of education from religion, most likely, are not the ones directly responsible for or actually carrying out the complement process. Note that the use of the infinitival complement instead of the gerundive one produces a clear shift in meaning in that in this case the subject of the matrix clause is clearly co-referent with the complement clause agent(s), as illustrated in (7):

(7) Chcemy całkowicie *oddzielić* edukację od religii.
   Lit. 'We want to completely separate education from religion.'

The impact of the subject (non)coreferentiality on the complement choices of *chcieć* becomes even more visible in cases where the complement agent is specified rather than implicit. Compare (8) and (9):

(8) Czego chcemy od rządu? *Podniesienia* zarobków i ograniczenia wzrostu cen.
   Lit. 'What do we want from the government? Raising the earnings and limiting the price increase.'

(9) Czego chcemy od rządu? *Podnieść zarobki i ograniczyć wzrost cen.
   Lit. 'What do we want from the government? To raise the earnings and limit the price increase.'

The above examples show that the gerundive construction is selected in situations when the main clause agent(s) is/are non-coreferent with the complement agent(s), with the latter typically remaining unspecified and not expressed linguistically. How-

ever, a closer look at the relevant usage patterns reveals that (non)coreferentiality is not an all–or-nothing affair. Consider the following examples:

(10) Putin *chce utworzenia* strefy wolnego handlu między UE a Rosją.
Lit. 'Putin wants creating a free-trade area between EU and Russia.'

(11) Turcja *chce wejścia* do Unii Europejskiej i akceptuje intensywny nadzór i liczne pytania o jej praktyki wewnętrzne.
Lit. 'Turkey wants entering the EU and accepts the intense supervision and numerous questions about its internal policies.'

In both these examples the matrix subject can be regarded as co-referent with the complement agent. Yet, this coreferentiality is only partial. That is, in (10) Putin, in most likelihood, is not going to be directly or personally involved in the process of establishing a free-trade zone. His role is more that of a controller of supervisor of the process being described than its direct participant. On the other hand, in (11) Turkey is not an independent or autonomous decision-maker regarding the process of EU enlargement – all member countries participate in decision making. What seems to follow from these examples is that (non)coreferentiality should be seen as a gradable category, with different degrees of the matrix subject's involvement in the complement process. That is, apart from strictly non co-referent uses, the gerundive complement can also be found in constructions with coreferent agents. Yet, in this latter case, the matrix subject's involvement in the complement process is subject to some restrictions or limitations in the sense that he/she is construed as one of the many agents responsible for the complement process. Note that in such situations the gerund can be replaced with the infinitive. Hence, one could also say:

(12) Putin *chce utworzyć* strefę wolnego handlu między UE a Rosją.
     Lit. 'Putin wants to create a free trade zone.'

(13) Turcja *chce wejść* do Unii Europejskiej […]
     Lit. 'Turkey wants to enter EU […]'

The difference between (10) and (12), on the one hand, and (11) and (13), on the other, is a subtle one, i.e. in the infinitival construction the matrix agent appears to be construed as the main effector of the complement process and therefore as being more in control of the process described in the complement clause than the subject of the corresponding gerundive constructions, whose involvement in the complement event is subject to some limitations, as argued above. The conclusion that follows from these considerations is that the (non)coreferentiality of the matrix and complement agents is a key factor in distinguishing between the gerundive and infinitival constructions, as they occur with *chcieć*. However, the distinction is a matter of degree rather than a clear-cut dichotomy. In order to substantiate this hypothesis with quantitative data, a corpus-based analysis has been performed on random samples of the two constructions under consideration, as extracted from Polish Web 2012 corpus (cf. section 2). All the occurrences have been coded for the variable of subject coreferentiality. Apart from the unambiguously coreferential and non-coreferential uses, the category of partial coreferentiality, as discussed above, has been distinguished. Also, some ambiguous cases have been found in the data, which could be hardly judged as being coreferent or non-coreferent from the limited amount of the context provided by the concordancer. The results of this analysis are shown below:

$\chi^2_{\text{Pearson}}(3) = 100.64, p = 1.13\text{e-}21, \widehat{V}_{\text{Cramer}} = 0.86, \text{CI}_{95\%} [0.70, 1.00], n_{\text{obs}} = 133$
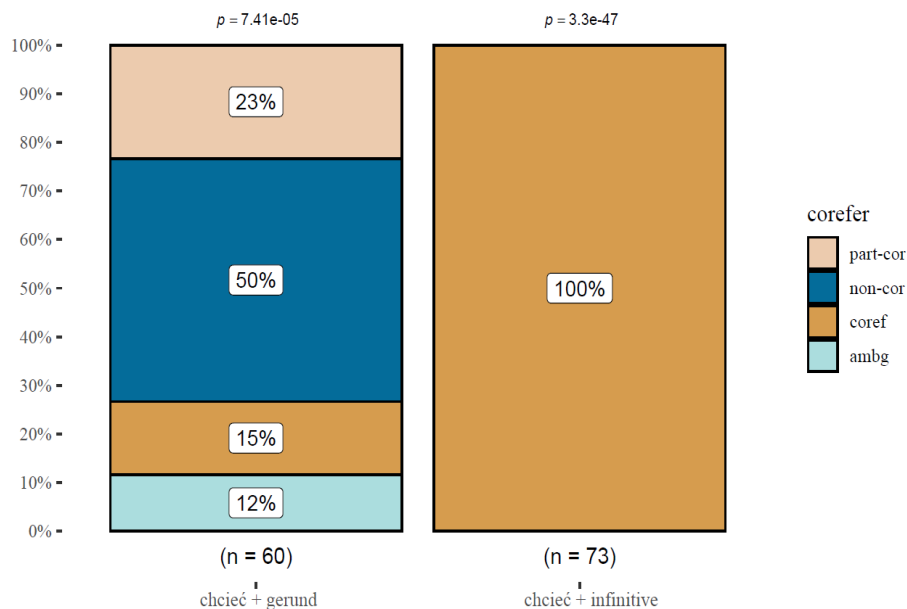
Figure 1

Coreferentiality of the subjects and
the choice of the complement type by *chcieć*

The analysis has confirmed the significance of the coreferentiality variable: while the infinitival complement is selected exclusively by unambiguously coreferent constructions, the gerundive construction clearly favours non-coreferent uses, with unambiguous cases accounting for 50 % and partially coreferent ones for 23 % of the sample being analyzed. Note that coreferent uses constitute merely 15 % of the gerundive sample. As expected, the Person's chi-squared test confirmed that these differences are highly significant: χ^2 (3) = 100.64, p = 1.13e-21. The overall effect is large: Cramer's V = 0.86.

4. Case study 2: *Zacząć* and its complements

As can be seen from Tables 1 and 2, there are three verbs of aspect that accept both types of non-finite complements: *zacząć*, *rozpocząć* and *zaprzestać*. The two nearly synonymous verbs of inception – *zacząć* and *rozpocząć* differ in their preferred complement types in that the former occurs more often with the infinitive, whereas the latter preferentially selects the gerundive complement. It is interesting to note at this juncture that *rozpoczynać* is semantically distinct from *zacząć* in that it denotes the onset of iterative, generic processes or events, whereas *zacząć* is used primarily in the context of specific, one-shot actions performed by a definite agent. Hence, while (14) below is a perfectly well-structured sentence, (15) is rather awkward, as 'rozpocząć pisanie' suggests a long term process rather than an instantaneous action. Hence, (16) is a more natural alternative to (15).

(14) Japończycy *rozpoczęli* bombardowanie strategicznych rejonów Singapuru.
Lit. 'The Japanese started bombarding the strategic regions of Singapore.'

(15) ?Wzięła do ręki długopis i rozpoczęła pisanie.
Lit. 'She picked up a pen and started writing.'

(16) Wzięła do ręki długopis i zaczęła pisać.
Lit. 'She picked up a pen and started to write.'

Also, it is worth pointing out here that *zacząć* is less restrictive and can be used as a (less formal) substitute for *rozpocząć*. Consider (17):

(17) Japończycy *zaczęli* bombardowanie strategicznych rejonów Singapuru.
Lit. 'The Japanese began bombarding the strategic regions of Singapore.'

This distribution appears to suggest that the gerundive complement tends to form a tighter relationship with the predicates that favour iterative, repetitive processes. A further confirmation of this hypothesis might be the fact that the gerund is clearly incompatible with bodily actions performed by specific agents on specific occasions, which can hardly be conceptualized in iterative terms. In such cases, the infinitive is the only possible choice:

(18) Kiedy usłyszał to, *zaczął płakanie/mówienie/ubieranie się […]
Lit. 'When he heard this, he started crying/talking/getting dressed.'

(19) Kiedy usłyszał to, zaczął płakać/mówić/ubierać się …
Lit. 'When he heard this, he started to cry/talk/get dressed […].'

Again, in order to verify this preliminary hypothesis a random sample of both constructions has been coded for the situation type denoted by the complement clause. This variable has been broken into three levels: single action, iterative action/ event, non-agentive process. The results are presented in Figure 2.

$$\chi^2_{\text{Pearson}}(2) = 20.13, p = 4.25\text{e-}05, \widehat{V}_{\text{Cramer}} = 0.32, \text{CI}_{95\%} [0.17, 1.00], n_{\text{obs}} = 182$$
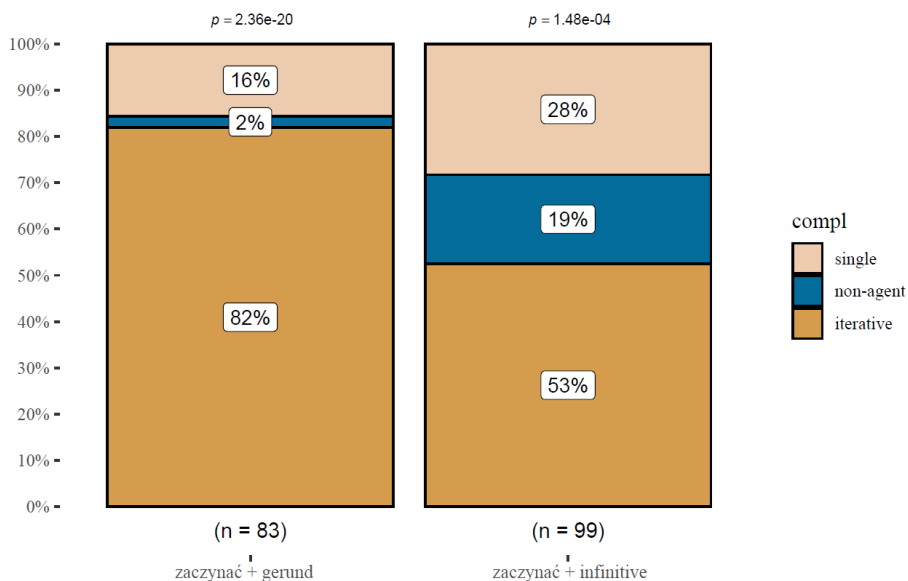
Figure 2
Complementation of *zaczynać* and subject types

As can be seen, both constructions are most often used to ex-press iterative events. However, the gerundive complement is clearly skewed towards iterative actions, whereas the infinitival construction shows a more even distribution of the different sit-uation types. Note also that non-agentive processes tend to be coded by the infinitival complement rather than the gerundive, which is illustrated with the following examples:

    (20) Jesienią zaczyna robić się spokojniej.
         Lit. 'In autumn it starts to get calmer.'

    (21) (…) 32 proc. Brytyjek poniżej 30 roku życia zaczyna siwieć.
         Lit. '32 per cent of British women under 30 starts to get grey.'

The Pearson's chi-squared test shows that there is a significant association between the type of situation coded by the comple-

ment clause and the type of the construction (χ^2(2, N = 182) =20.1; p < 0,001). As indicated by a post-hoc test applied to these data, the significant result is due to two types of association, i.e. between iterative vs. non-agentive events (padj <0.001; V = 0.34), on the one hand, and iterative vs. single events (padj =0.015; V = 0.22), on the other. At the same time no significant relationship has been observed between iterative vs. single events (padj = 0.106; V = 0.25). However, the overall effect is only moderate (V = 0.33).

Another variable that has been subjected to an empirical test is the type of the subject. The qualitative inspection of the concordance lines has revealed a strong presence of impersonal subjects in the gerundive sample. In particular, it has been observed that the gerundive complement tends to co-occur with the "się" impersonal reflexive forms, as illustrated below. Note that the infinitive is unacceptable in this context:

(22) W końcu zaczęło się odcinanie kuponów, zresztą bardzo dochodowe.
Lit. 'Finally, clipping off coupons began, [...].'

(23) *Zaczęło się odcinać kupony, zresztą bardzo dochodowe.
Lit. 'Coupons began to be clipped off [...].'

(24) Tak zaczęły się aresztowania i skazania winnych bratobójczej śmierci.
Lit. 'It was how the arrests and convictions of the guilty of fratricidal death began.'

(25) *Tak zaczęły się aresztować i skazać winnych bratobójczej śmierci.
Lit. 'It was how to arrest and convict the guilty of fratricidal death began.'

It is interesting to note here that the substitution of personal subjects for the impersonal ones makes the infinitive fully acceptable.
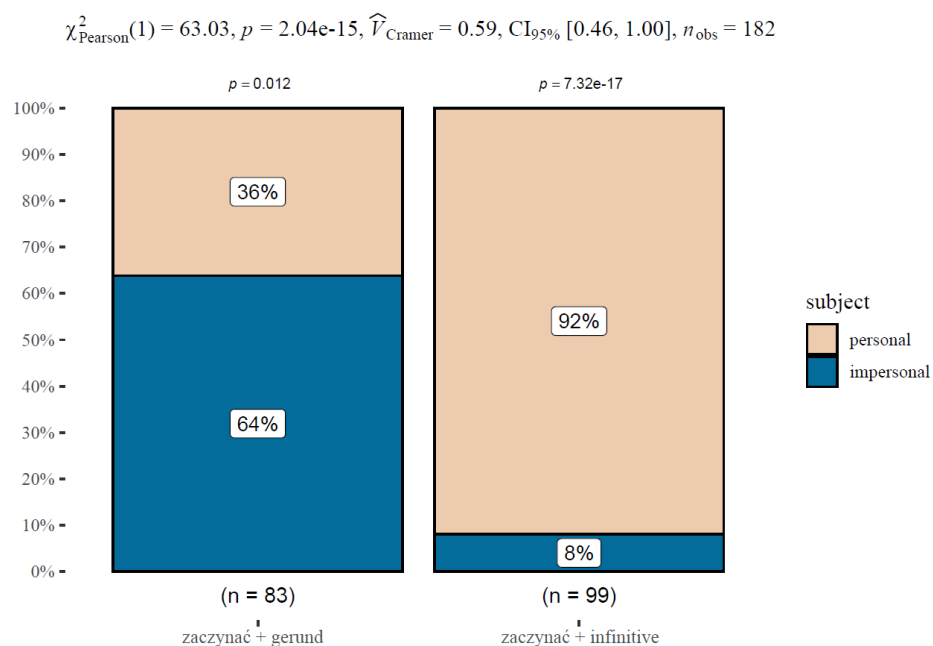
(26) W końcu zaczęła odcinać kupony.
     Lit. 'Finally, she started to clip off the coupons.'

(27) Tak zaczęli aresztować i skazywać winnych bratobójczej
     śmierci.
     Lit. 'It was how they started to arrest and convict the guilty of
     fratricidal death.'

Given these preliminary observations, the same sample as be-
fore has been coded for the subject type variable, which has
been broken into two levels: impersonal constructions (i.e. with
unexpressed agents) vs. personal constructions (i.e. with ex-
pressed agents). The results are presented in Figure 3:

$\chi^2_{\text{Pearson}}(1) = 63.03, p = 2.04\text{e-}15, \widehat{V}_{\text{Cramer}} = 0.59, \text{CI}_{95\%} [0.46, 1.00], n_{\text{obs}} = 182$

Figure 3
Personal vs. impersonal subjects in *zaczynać* constructions

As can be seen, the infinitival construction clearly favours personal subjects (92 %), with impersonal forms accounting for merely 8 % of the sample. In the case of the gerundive constructions, these proportions are reversed – the impersonal uses clearly dominate, accounting for 65 % of the sample. The chi-squared test confirmed that there is a significant association between the subject type and the construction (χ^2(1, N = 182)=60.55; p < 0.001). The overall effect is large rφ = 0,59.

To sum up, although the two constructions under consideration overlap to some extent, some clear preferences (and dispreferences) can be observed in their usage patterns in terms of the type of the situation coded by the complement clause and the type of the subject referent in the matrix clause. In both cases there is a statistically highly significant correlation between the constructional choices and the variables in question. However, as indicated by the respective Cramer's V values, the association is much stronger in the case of the subject type than the situation-type variable.
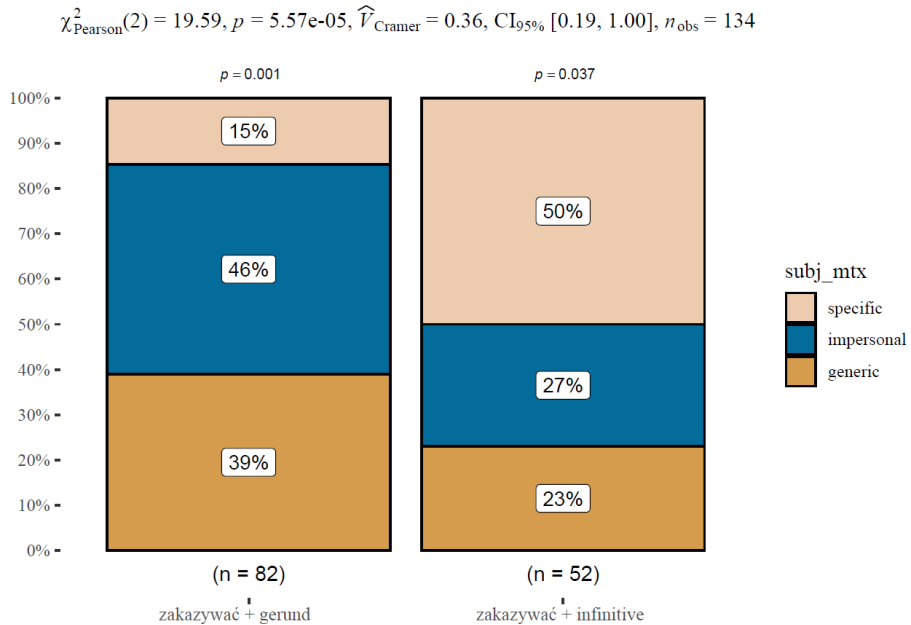
5. Case study 3: *Zakazać* (forbid) and its complements

The notion of (im)personality also appears to play a significant role in the selection of the complement type by *zakazać*. Consideration of the relevant concordance lines has revealed that the infinitive is selected more often if the main clause subject is specific and also when the lower clause agent is given in the sentence and has a specific referent. The gerundive complement, on the other hand, favours generic subjects in both clauses and is also selected when the agents of the complement clause are unspecified and unexpressed. The following examples illustrate these two main patterns.

(28) Ja mu nie zakazuję mówić w ojczystym języku [...].
     Lit. 'I don't forbid him to speak his native tongue.'

(29) W USA zakazano kiedyś picia alkoholu, a pił prawie każdy [...]
    Lit. 'In the USA drinking alcohol was once forbidden, yet eve-
    ryone was drinking.'

As in the two previous studies, these preliminary observations
have been subjected to quantitative verification. That is, a sam-
ple of the two constructions has been coded for the type of ma-
trix subject and the type of the complement subject. The former
has been broken into three levels: impersonal, generic, specific,
and the latter into: not expressed, generic and specific. Figure 4
presents the distribution of the matrix subjects.

$$\chi^2_{\text{Pearson}}(2) = 19.59, p = 5.57\text{e-}05, \widehat{V}_{\text{Cramer}} = 0.36, \text{CI}_{95\%}\ [0.19, 1.00], n_{\text{obs}} = 134$$
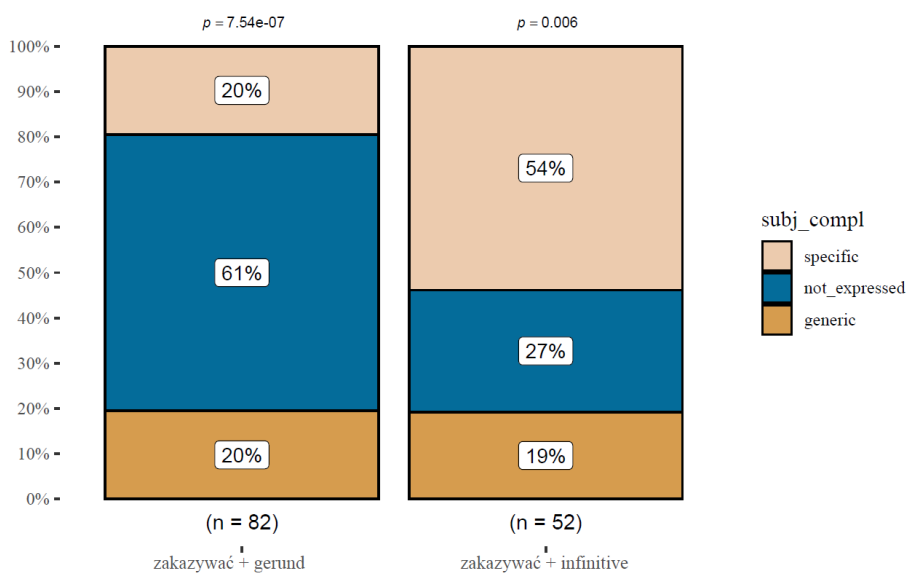

Figure 4
Type of the matrix subject in *zakazywać* constructions

As these frequencies indicate, the gerundive construction clear-
ly favours impersonal and generic subjects in the main clause,
which account for 46 % and 39 % of the sample, respectively,

with specific subjects being rather marginal (15 %). The infinitival construction, on the other hand, prioritizes subjects with specific referents (50 %) over the two other types. The chi squared test shows that the type of the matrix clause subject is highly significantly correlated with the choice of the complement construction (χ^2(2, N = 134)=19.59; p < 0.001). However, the overall effect is moderate $r\varphi$ = 0.36.

Let us now consider the type of the subject in the complement clause. Also here three levels have been distinguished: subjects with specific (definite) referent, subjects with generic referents and subjects that are not coded. The quantitative analysis has revealed the distribution presented in Figure 5.

$$\chi^2_{\text{Pearson}}(2) = 19.15, p = 6.94\text{e-}05, \widehat{V}_{\text{Cramer}} = 0.36, \text{CI}_{95\%}\ [0.19, 1.00], n_{\text{obs}} = 134$$

Figure 5
Type of complement subjects in *zakazywać* constructions

As can be seen, the distributional patterns of complement subjects are similar to the ones observed in the main clause: in the infinitival construction, specific subjects prevail (59 %) over the

two other types, whereas the gerundive construction clearly favours impersonal/non-expressed subjects (61 %). The chi-squared test confirmed that the there is a highly significant correlation between the choice of the complement construction and the type of complement subject (χ^2(2, N = 134)=19.1; p < 0.001). The overall effect size is moderate rφ = 0.36.

In sum, the obtained data allow to conclude that *zakazać* + gerund significantly favours impersonal or generic contexts of use, whereas *zakazać* + infinitive is significantly skewed towards specific situations, i.e. actions performed by specific agents. However, given that the correlation between the two variables is only moderate, some overlaps between the two constructions can be expected.

6. Reduced-increased involvement and dispersal of control

The three pairs of constructions considered in this paper denote distinct kinds of processes, which is hardly surprising given the semantic distinctiveness of the matrix predicates. Yet, the distributional data presented in the three case studies show a striking consistency in the way the two complement constructions under consideration portray the relationship between the main clause agent and the complement process. The pattern that emerges from the data is that the choice of one or the other complement type has to do with the degree of the matrix agent's control over the process described in the complement clause. As has been seen, the gerundive construction significantly prefers non-coreferent uses (cf. *chcieć*) as well as unspecified or generic subjects (cf. *zacząć*, *zakazać*) over the coreferent and definite/specific subjects. It should be clear that in non-coreferential constructions it is the complement subject, not the matrix subject that is the 'doer' of the complement action. Hence, the latter is only indirectly involved in the complement process and therefore has only limited influence on its outcome. It follows from this that the gerundive construction with *chcieć* profiles

the notion of reduced or limited control that the main clause agent has over the complement process, whereas the corresponding infinitival construction portrays the matrix agent as the main doer and hence the one having the complement action directly under his/-her control.

Also, we have seen that the notion of control is a matter of degree rather than being an all or nothing phenomenon in that the subject referent can be construed as one of the agents responsible for the complement process (cf. examples 10, 11). This situation involves what might be described as a 'dispersal or control' in the sense that the responsibility for the successful completion of the action is spread over different actors, rather than being exclusively in the hands of the matrix subject referent. As we have seen, in such intermediate cases *chcieć* can be used with both types of complements more or less interchangeably. Yet, the choice of the infinitive or the gerund involves a subtle shift in perspective, which proceeds in the direction predicted by this analysis, i.e. the infinitival constructions present the subject as being more in control, whereas the gerundive ones profile the notion of reduced controllability on the part of the matrix agent.

The type of subject has also be found to be a significant factor in the complement choices of *zaczynać* and *zakazywać*. These choices are significantly influenced by whether the agents are specified or unspecified. As has been seen, the gerundive constructions clearly favour unspecified/impersonal subjects, while the infinitival ones prefer definite/specified agents. This distribution appears to confirm the control/ involvement theory as formulated above in that the unspecified agents can hardly be construed as having any control or being actively involved into a process. In other words, if the agent is unknown, or unexpressed, his/her causal influence is either irrelevant or de-focused. If, on the other hand, the agents are defined and specific, the issue of control becomes much more relevant and figures more prominently in the description of the scene.

It is advisable to point out here that *zakazać* is distinct from the two other predicates in that it obligatorily requires that the subject of the main clause is non co-referent with the subject of the complement clause. This, in turn, appears to explain why *zakazać* is more significantly associated with the gerundive complement, while the two other predicates favour the infinitival complement. Given that non-coreferentiality entails that the matrix subject is less in control or less directly involved in the complement action (for the simple reason that we have less control over the actions of others than over our own action), this distribution appears to be motivated by the semantics of the gerundive complement, i.e. by its tendency to portray the complement process as being somewhat 'detached' from the influence of the main clause agent(s).

It might follow from the above that the theory of reduced-increased controllability and/or dispersal of control, as advocated in this paper, is applicable only to agentive events, which presuppose the notion of control. The question that arises at this point is how this theory relates to non-agentive events, which are coded by the infinitival complement rather the gerundive one, although they are not controllable in the strict sense of the word (cf. examples 20, 21). Yet, this distribution is not contradictory if one considers that these non-volitional process (e.g. one's hair going grey or getting old), portray the subject referent as an experiencer of the process described in the complement clause, or a thing that is directly affected by the complement process. Hence, also here the subject referent is construed as directly involved in the complement process, although it is 'experiential' rather than causative involvement. This, in turn, is in line with the theory proposed in this paper in that the degree of one's involvement in a process (in the sense of being more or less affected by it) and the degree of one's causal influence on an action appear to represent two sides of the same coin.

Finally, it must be reiterated that most of the semantic contrasts discussed in this paper are gradient rather than absolute, which is consistent with the generally accepted view that lin-

guistic choices are probabilistic in nature rather than categorical. Nevertheless the fact remains that certain construction types are strongly and significantly preferred in certain types of contexts. The distributional preferences (or dispreferences), as discussed in this paper, provide evidence for another widely accepted view, namely that full synonymy is a rare occurrence and a change in form always spells a difference in meaning, as famously argued by Bolinger (1968).

7. Concluding remarks

The corpus-based, bottom-up approach employed in this study has allowed to determine the linguistic factors that significantly affect the choice of the infinitival or gerundive complement in cases where both types are accepted by a given matrix predicate. It should be clear that one would not be able to detect all these significant patterns of use with purely introspective or intuition-based methods. Thus, the role of corpus-based methods in analysing closely related or nearly-synonymous constructions is not to be overestimated. However, it must be emphasized at this point that the identification of significant patterns of use has not been the ultimate goal of this research. Distributional patterns have been taken here to represent or to be indicative of semantic patterns, which is a particular case of the distributional hypothesis. The qualitative interpretation of the frequency data has confirmed this line of thinking, which is to say that some general principles governing and motivating the infinitive-gerund alternations have emerged from the data and have been generalized as a *reduced-increased involvement theory*. However, given the limited scope of the present study, this theory should be viewed as a working theory rather than a final or fully confirmed one. There are still many questions that need to be addressed in future research. First of all, it remains to be seen how this account relates to constructions governed by other verbs that participate in infinitive-gerund alternations – a task that could not been undertaken in this paper due to spa-

ce limitations. Hence, possibly other linguistic/contextual factors will be discovered that significantly affect the choices between the two constructional variants under analysis. The next worthwhile step seems to be the application of multivariate statistics in order to determine the predictive power of particular variables being discussed. Also, a comparison of the findings presented in this study with data from other languages might provide a worthwhile insight into the cross-linguistic validity of the theory proposed in this study.

References

Bolinger, Dwight (1968). "Entailment and the meaning of structures". *Glossa* 2: 119–127.

Bresnan, Joan (1979). *Theory of Complementation in English Syntax*. New York: Garland.

Deshors, Sandra C. (2015). "A constructionist approach to gerundial and infinitival verb complementation patterns in native and Hong Kong English varieties". *English Text Construction* 8/2: 207–235.

Deshors, Sandra C., Stefan Th. Gries (2016). "Profiling verb complementation constructions across New Englishes: A two-step random forest analysis of *ing* vs. *to* complements". *International Journal of Corpus Linguistics* 21/2: 192–218.

Dirven, René (1989). "A Cognitive Perspective on Complementation". In: Dany Jaspers, Wim Klooster, Yvan Putseys, Pieter Seuren (eds.). *Sentential Complementation and the Lexicon*. Dordrecht: Foris Publications, 113–139.

Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele (2006). *Constructions at Work*. Oxford: Oxford University Press.

Givón, Talmy (1980). "The binding hierarchy and the typology of complements". *Studies in Language* 4: 333–377.

Givón, Talmy (2001). *Syntax: An Introduction*, Vol. 2. Amsterdam: John Benjamins.

Gries, Stefan Th., Anatol Stefanowitsch (2004). "Extending collostructional analysis: A corpus-based perspectives on 'alternations'". *International Journal of Corpus Linguistics* 9/1: 97–129.

Gries, Stefan Th. (2017). "Syntactic alternation research. Taking stock and some suggestions for the future". *Belgian Journal of Linguistics* 31: 8–29.

Harris, Zellig S. (1954). "Distributional Structure". *WORD* 10/2-3: 146-162.

Kaleta, Agnieszka (2012). "The English gerund vs. the to-infinitive: the case of aspectual constructions". *Selected Papers from UK-CLA Meetings,* Volume 1: 323–34.

Kaleta, Agnieszka (2014). *English Sentential Complementation: A Usage-based Approach*. Piotrków Tryb.: Naukowe Wydawnictwo Piotrkowskie.

Krzeszowski, Tomasz P. (1990). *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin – New York: Mouton De Gruyter.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar* Vol. 1. Stanford: Stanford University Press.

Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar* Vol. 2. Stanford: Stanford University Press.

Langacker. Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Noonan, Michael (1985). "Complementation". In: Timothy Shopen (ed.). *Complex Constructions*. Cambridge: Cambridge University Press, 42-140.

Rudanko, Juhani (2017). *Infinitives and Gerunds in Recent English: Studies of Non-Finite Complements with Data from Large Corpora*. Basingstoke: Palgrave Macmillan.

Rudanko, Juhani (2011). *Changes in Complementation in British and American English: Corpus-based Studies on Non-finite Complements in Recent English*. Basingstoke: Palgrave Macmillan.

Rudanko, Juhani (2021). *Infinitival vs Gerundial Complementation with Afraid, Accustomed, and Prone. Multivariate Corpus Studies*. Basingstoke: Palgrave Macmillan.

Smith, Michael B. (2008). "The semantics of complementation in English: A cognitive semantic account of two English complement constructions". *Language Sciences* 31: 360–388.

Smith, Michael B., Joyce Escobedo (2001). "The semantics of *to*-infinitival vs -*ing* verb complement constructions in English". In: Mary Andronies, Christopher Ball, Heidi Elston, Sylvain Neuvel (eds.).

Proceedings from the Main Session in the Chicago Linguistic Society's Thirty-seventh Meeting. Chicago: Chicago Linguistic Society, 549–64.

Taylor, John, Rene Dirven (1991). *Complementation.* Duisburg: LAUD Linguistics Agency, University of Duisburg.

Stefanowitsch, Anatol, Stefan Th. Gries (2003). "Collostructions: Investigating the interaction between words and constructions" *International Journal of Corpus Linguistics* 8/2: 209-43.

Wierzbicka, Anna (1988). *The Semantics of Grammar.* Amsterdam: John Benjamins.

Agnieszka Kaleta
ORCID iD: 0000-0002-6272-7163
Institute of Literature and Linguistics
Jan Kochanowski University
ul. Słowackiego 114/118
97-300 Piotrków Trybunalski
Poland
agnieszka.kaleta@ujk.edu.pl