The English Gerund vs. The to-infinitive: The Case of Aspectual Constructions

Selected Papers from UK-CLA Meetings
http://uk-cla.org.uk/proceedings
Vol 1: 323 – 341

AGNIESZKA KALETA
Jan Kochanowski University
agakaleta@poczta.onet.pl

The present paper is concerned with the complementation patterns of four aspectual verbs, i.e. begin, start, continue and cease, each of which co-occurs with two types of non-finite complementizers – the gerund and the to-infinitive. The paper analyses the distributional properties of each pair of the aspectual constructions (e.g. begin to do/ begin doing) with the corpus-based method known as a distinctive collexeme analysis – a method geared specifically to investigating the interaction between syntactic constructions and lexemes which fill those constructions. The analysis shows that the pairs of constructions with the same matrix verb do not represent arbitrary alternations but are semantically motivated. The main aim of this contribution is to shed some new light on the schematic semantics of the gerundive and infinitival complements and thereby account for the motivated nature of their distribution.

Keywords: aspectual verbs, complementation, gerund, to-infinitive, distinctive collexeme analysis

1. Introduction[i]

The present paper takes up the issue of the semantic contrast between two English non-finite complementizers – the gerund and the to-infinitive. It has been long argued in the cognitively and functionally oriented literature that these two categories have the status of constructions, that is form-meaning pairings, the distribution of which is semantically motivated (cf. Bolinger 1968; Wierzbicka 1988; Dirven 1989; Duffley 2006). However, the exact nature of the semantic contrast between these two constructions has not been conclusively established, at least not in a way that could be subject to rigorous empirical verification.

In this paper I make an attempt to provide an empirically and cognitively plausible account of the two English complement constructions. More specifically, I take a look at the gerundive and infinitival complementizers as they co-occur with verbs of aspect: i.e. *begin, start, continue, cease* (henceforth aspectual constructions). Each of these verbs can take both complementizers without any obvious semantic consequences. However, the lack of apparent semantic differences does not entail semantic equivalence. On the contrary, given the common assumption that a change of form almost

invariably implies a change of meaning (e.g. Bolinger 1968) the pairs of constructions with the same matrix verb and two different complementizers (e.g. begin to do/begin doing) do not represent arbitrary alternations, but are semantically motivated. The main aim of the present paper is to substantiate this thesis with empirical evidence. More specifically, I trace the distributional properties of the aspectual constructions as defined above utilizing the so called *distinctive collexeme analysis* – a method geared specifically to investigating the interaction between syntactic constructions and lexemes which fill those constructions. (cf. section 2). However, tracing and recording the distributional characteristics of the nearly synonymous aspectual constructions is not an end in itself. It is a means to a more fundamental goal, namely uncovering the underlying semantic factors which motivate the distribution of the gerundive and the to-infinitival complementizers and thereby delineating the nature of the semantic contrast between these two categories.[ii]

The paper is organized as follows. The next section briefly characterizes distinctive collexeme analysis as introduced and popularized in a series of papers by Anatol Stefanowitsch and Stefan Th Gries (cf. Stefanowitsch and Gries 2003; Gries and Stefanowitsch 2004, 2010). Section (3) provides the general frequency data – the raw frequencies of particular constructions and the number of verb types filling the constructional slots. Section (4) presents the results of the distinctive collexeme analysis as applied to the infinitival and gerundive constructions of aspect, which will be mnemonically referred to as: BEGIN_TO, BEGIN_ING, SART_TO, START_ING, CONTINUE_TO, CEASE_TO, CEASE_ING. Section (5) discusses the distributional data presented in section (4) in the light of the general assumption that the distribution of the aspectual constructions is not random or arbitrary but semantically motivated.

2. Distinctive Collexeme Analysis

Distinctive collexeme analysis is a corpus-based method of syntactic analysis, which is an extension of collostructional analysis – a method concerned with investigating the relationship between words and constructions associated with them (rather than focusing on the relationship between different lexemes as in a typical collocational analysis). More specifically, collostructional analysis (a blend of construction and collocational) aims at measuring the degree of attraction or repulsion that words exhibit to constructions.[iii] The main idea behind this approach is as follows:

> (…) a word may occur in a construction if it is *semantically compatible* with the meaning of the construction (or, more precisely, with the meaning assigned by the construction to the particular slot in which the word appears). For example, the verb *give* may occur in the ditransitive construction because the verb and construction have the same meaning ('sb transfers sth to sb'). (Stefanowitsch and Gries 2003: 213)

Distinctive collexeme analysis is specifically geared towards investigating semantic properties of pairs of semantically similar grammatical constructions. Its main aim is to identify lexemes that exhibit a significantly

higher preference for one member of the pair than for the other. The method compares the frequencies of words that occur in one constructional slot to their frequencies in a corresponding slot in the nearly equivalent construction, and thus makes it possible to identify subtle distributional differences between the members of such a pair. As claimed by the authors distinctive collexeme analysis has two main applications:

> first, to increase the adequacy of grammatical description by providing an objective way of identifying the meaning of a grammatical construction [...] second, to provide data for linguistic theory-building. (Stefanowitsch and Gries 2003: 209)

In this paper the collocational preferences of the gerundive and infinitival complementizer are taken to reflect the semantic differences between these two constructions. Four different pairs of aspectual constructions as defined in the introduction are examined with a view to identifying the lexemes that best distinguish between the two non-finite complementizers under investigation.

It should be added here that the strength of association between the constructions in question and the lexemes that fill them are expressed in terms of p-values yielded by the Fisher Exact test (cf. ~~Pednderson~~ Pedersen 1996) recommended by the authors as most reliable in this kind of analysis.[iv] The analyses presented in the following sections are based on the data extracted from the British National Corpus (henceforth BNC).

3. Aspectual Constructions: General Frequencies

In this section I look at the raw frequencies of the aspectual constructions as drawn from the BNC. The relevant data are included in the second column of the table below. The third column presents the total frequencies of the infinitival and gerundive constructions for particular verbs of aspect.

Construction	Freq.	total
BEGIN_TO	19700	22580
BEGIN_ING	2880	
START_TO	6356	14125
START_ING	7769	
CONTINUE _TO	12024	13070
CONTINUE_ING	1046	
CEASE_TO	1590	1792
CEASE_ING	202	

Figure 1. Raw frequencies of aspectual constructions in the BNC

As shown in Figure 1 there is a marked quantitative discrepancy between the infinitival and the gerundive constructions, which is to say that the former significantly outnumber the latter. The only clear exception to this tendency are the START constructions, in which case the proportions between the two aspectualizers are much more even, with a slight leaning towards the gerundive form. It should be also observed that the totals for particular matrix verbs also vary significantly. The BEGIN constructions are by far most frequent with over 22 000 occurrences and are followed by START and CONTINUE constructions, which account for approximately the same number of occurrences - circ. 14 000 and 13 000 respectively.

The non-finite CEASE forms lag far behind with the total number of occurrences below 2000.

In the table below I present the frequencies of particular verb types filing the complement slots of the aspectual constructions. The table shows the number of different verb types that co-occur with one or the other complementizer, the number of verb types shared by the two complementizers and the totals for each construction (with corresponding percentages in the brackets).

	to-infinitive only collexemes	**ING_only collexemes**	**common collexemes**	**TOTAL**
BEGIN	1416 (61%)	222 (10%)	665 (29%)	2303 (100%)
START	567 (32%)	538 (30%)	663 (38%)	1768 (100%)
CONTINUE	1155 (75%)	99 (7%)	276 (18%)	1530 (100%)
CEASE	225 (71%)	53 (17%)	39 (12%)	317 (100%)

Figure 2. Aspectual constructions by the number of different verb types.

As can be seen there is a notable correlation between Figure 1 and 2 in terms of the quantitative discrepancies between the infinitival and the gerundive constructions. This is to say that in the case of the BEGIN, CONTINUE and CEASE constructions the infinitival slot attracts many more different verb types than the gerundive one. In the case of the START constructions these numbers are almost equally spread between the two slots. What should be emphasized here is also the fact that in each of the constructions there is a significant proportion of verbs shared by the two complementizers.

The main question to be asked now is whether these general frequencies have any qualitative basis. In other words, whether the significant quantitative disproportions between the infinitival and gerundive complementizers of the four aspectual verbs under considerations are accompanied by any semantic differences or contrasts. This is a question that will be addressed in the next section in a series of studies utilizing distinctive collexeme analysis.

4. Aspectual Constructions and their Non-finite Complementation: Distinctive Collexeme Analysis

4.1 Introduction

The analysis has been carried out in a few stages consisting of the following procedures:

- The extraction of all the aspectual constructions as defined in the Introduction from the BNC corpus. This has been performed with the Sketch Engine software.[v] The query has been specified in such a ways as to extract all the verbs occurring in the infinitival and gerundive slots following the four matrix verbs: begin, start, continue, cease.
- Statistical evaluation of the strength of association between the four matrix verbs and their collexemes with Fisher Exact test.
- Semantic analysis of the data yielded by the collostructional analysis. In this analysis the most significant collexemes of the i~~nfinival~~ infinitival and gerundive constructions are taken to reflect the prototypical semantic properties of the two complementizers.

For reasons of space, I present and discuss only a selection of the most significant collexemes (i.e. 15-20 top lexemes), that is those which show the greatest strength of association with the infinitival and gerundive slots, as expressed in terms of the p-values yielded by the Fisher Exact test.[vi]

4.2 BEGIN_TO vs. BEGIN_ING

Here are the results of the distinctive collexeme analysis for the two BEGIN constructions:

collexeme	freq. in begin_to	freq. in begin_ing	Fisher Yeates
be	667	4	2,19E-34
feel	572	4	7,20E-29
wonder	251	4	6,24E-11
see	315	9	2,04E-10
grow	136	1	1,59E-07
fall	172	4	7,60E-07
show	180	8	7,00E-05
find	103	2	8,02E-05
think	375	29	2,11E-04
look	346	26	2,31E-04
enjoy	77	1	2,99E-04
appear	207	14	1,44E-03
cry	145	9	4,01E-03

sound	54	1	5,03E-03
go	105	7	2,05E-02
break	67	4	4,30E-02
sink	35	1	4,64E-02
slow	35	1	4,64E-02
form	73	5	5,73E-02
laugh	108	9	6,02E-02

Table 1. Top 20 collexemes distinctive for BEGIN_TO

collexeme	**Freq. in begin_to**	**Freq. in begin_ing**	**Fisher Yeates**
shipping	5	53	5,45E-42
working	115	70	3,34E-18
trading	2	16	5,40E-13
writing	86	49	2,19E-12
collecting	15	23	5,28E-12
training	5	17	8,03E-12
publishing	11	19	1,24E-10
marketing	4	14	5,25E-10
selling	17	18	3,80E-08
filming	2	10	5,67E-08
firing	4	11	1,16E-07
operating	136	50	1,51E-07
talking	25	20	1,51E-07
withdrawing	14	15	4,78E-07
teaching	23	18	7,92E-07
trying	18	16	1,03E-06
distributing	2	7	1,53E-05
delivering	2	7	1,53E-05
using	73	29	1,86E-05
experimenting	16	13	2,07E-05

Table 2. Top 20 collexemes distinctive for BEGIN_ING

To begin with the BEGIN constructions – a comparison of the collexemes in Tables 1 and 2 points to some essential characteristics and contrasts between the two constructions. In most general terms, the dividing line between these two forms lies between stative and non-agentive processes, on the one hand, and agentive events (action verbs), on the other. As can be seen the top position on the collexeme list of the BEGIN_TO construction is occupied by the general stative verb 'be', which is followed by a range of items referring to cognitive and emotional states – 'feel', 'wonder', 'think', 'enjoy', 'see'. There is also a set of high frequency verbs encoding non-agentive processes of change i.e. 'grow', 'fall', 'sink'. The list in Table 1 clearly contrasts with the data included in Table 2, which contains lexemes denoting various agentive processes. What is noteworthy here is that most of these processes have iterative nature implying 'a series of events', that is repeated episodes of doing something. For instance, 'begin shipping' implies a series of shipments rather than a single episode of transporting something. The same can be said about many other items such as 'trading/collecting/publishing/selling/ distributing/delivering/experimenting/trying', all of which imply long-term serial events, or, to put it differently, a sequence of events.

Let us now take a closer look at the nearly synonymous START constructions to see if they bear any similarities to what has been observed for the BEGIN constructions.

4.3 START_TO vs. START_ING

The results of the relevant collostructional analysis have been presented in Tables 3 and 4 below:

collexeme	freq. in start_to	freq. in start_ing	Fisher Yeates
feel	123	30	6,54E-022
emerge	31	1	5,66E-011
appear	44	8	4,52E-010
become	48	12	3,16E-009
rise	37	7	1,62E-008
fall	59	27	8,96E-007
climb	32	8	1,43E-006
rain	38	12	1,54E-006
grow	53	25	5,01E-006
show	45	19	6,34E-006
change	31	9	7,12E-006
turn	47	21	8,35E-006
get	225	201	9,72E-006
return	16	1	1,17E-005
pick	28	8	1,78E-005

fade	15	1	2,59E-005
move	84	57	3,24E-005
relax	14	1	5,72E-005
crumble	14	1	5,72E-005
shake	32	13	9,83E-005

Table 3. Top 20 collexemes distinctive

collexeme	**Freq. in start_to**	**Freq. in start_ing**	**Fisher Yeates**
talking	52	260	5,00E-023
shipping	5	51	3,08E-008
doing	63	181	3,24E-008
training	4	42	4,58E-007
thinking	79	195	1,55E-006
shooting	3	36	1,66E-006
trying	13	60	6,19E-006
calling	6	41	1,03E-005
playing	51	133	2,00E-005
paying	15	56	1,21E-004
shouting	17	60	1,32E-004
saving	1	20	1,45E-004
dancing	4	28	2,57E-004
selling	13	49	2,93E-004
manufacturing	0	13	7,40E-004
yelling	0	13	7,40E-004
planning	11	41	1,00E-003
asking	24	66	1,28E-003
painting	4	24	1,39E-003
throwing	8	33	1,65E-003

Table 4. Top 20 collexemes distinctive for START_TO for START_ING

Even a cursory look at the data in question reveals that there is an analogy between these two sets of data. Most striking is the analogy between the gerundive complementizers of the two constructions in that they both show a very distinct tendency to attract agentive verbs.

There are a few items shared by the gerundive complementizers of BEGIN and START i.e. 'shipping', 'talking', 'training', 'trying' and 'selling'. This overlap constitutes one fourth of the top 20 collexemes of these two constructions. What should be also pointed out here is that both constructions tend to construe the events they encode as iterative, that is as a set of repeated processes. Particularly noteworthy is here the lexeme 'shipping', which inevitably denotes a series of different acts of transportation. Although the lexeme in question does not figure among the most frequent verbs in the English language, it ranks, respectively, as the most and the second most significant collexeme of the BEGIN and START constructions, which in turn makes it highly representative from the point of view of the prototypical semantic structure of the gerundive complementizer.

As far as the infinitival collexemes are concerned, we also find some points of convergence between the BEGIN and START constructions. The shared collexemes include the emotion verb 'feel' - highly distinctive for both constructions, followed by 'appear', 'fall', 'grow', and 'slow'. As in the case of the gerundive construction this overlap constitutes one fourth of the top 20 collexemes considered in this paper. What should be pointed out here is that among the most significant collexemes of the infinitival START construction, there is a substantial proportion of verbs referring to processes of change of state or transitions from one state of affairs into another state of affairs i.e. 'become', 'rise', 'fall', 'grow', 'change', 'turn', 'get', 'fade, 'crumble'.

4.4 CONTINUE_TO vs. CONTINUE_ING

Moving now on to the CONTINUE constructions, we also find some essential qualitative differences between the two non-finite complementizers, as revealed by the following data:

collexeme	freq. in continue_to	freq. in continue_ing	Fisher Yeates
be	1791	10	8,16E-056
have	189	3	8,27E-005
exist	70	0	2,71E-003
press	67	0	3,49E-003
hold	89	1	4,46E-003
rise	164	5	5,24E-003
support	124	3	6,60E-003
show	75	1	1,26E-002
attract	51	0	1,35E-002
grow	241	11	1,39E-002
fall	93	2	1,44E-002
increase	109	3	1,68E-002
believe	48	0	1,74E-002

dominate	47	0	1,90E-002
maintain	43	0	2,66E-002
flourish	38	0	4,06E-002
receive	79	2	3,57E-002
stare	56	1	4,91E-002
apply	35	0	5,23E-002
monitor	34	0	5,69E-002

Top 20 collexemes distinctive for CONTINUE_TO

collexeme	**Freq. in continue_to**	**Freq. in continue_ing**	**Fisher Yeates**
working	210	84	7,55E-026
talking	25	22	1,54E-012
walking	14	18	2,94E-012
reading	19	16	2,73E-009
rushing	127	36	2,01E-008
funding	10	11	1,50E-007
writing	41	18	4,94E-007
trading	34	16	1,04E-006
fishing	2	6	6,65E-006
knitting	9	8	2,21E-005
racing	2	5	6,23E-005
living	120	26	1,00E-004
collecting	6	6	1,65E-004
typing	1	4	1,98E-004
caring	11	7	3,16E-004
travelling	0	3	5,26E-004
breastfeeding	0	3	5,26E-004
financing	5	5	6,08E-004
fighting	41	12	8,64E-004
searching	14	7	9,31E-004

Table 6. Top 20 collexemes distinctive for CONTINUE_ING

Table 5 shows that the lexeme ‘be’ is by far most significant for the infinitival construction. As can be seen, there is a very substantial quantitative difference between this most strongly attracted item and all the other ones that follow it.

For instance, the next two items, that is the stative verbs 'have' and 'exist' are circ. 10 and 3 times respectively less significant for this construction than 'be'. However, together these three top collexemes indicate that the infinitival construction has a marked tendency to encode the continuation of stable (stative) configurations. Among the other collexemes can be found a set of verbs denoting non-agentive processes of change, or transition from one state of affairs into another state of affairs i.e. 'grow', 'fall', 'increase', 'flourish'. In addition, there is a single cognition verb 'believe'. The remaining items are more diversified in terms of their semantic content, yet most of them refer to agentive events of some sort (e.g. 'support', 'maintain', 'monitor', 'apply'). This, in turn, contrasts with our earlier observations that the infinitival construction encodes non-agentive processes. However, there seems to be an essential ontological difference between the agentive events coded by the infinitival construction and the gerundive one, as listed in Table 6. What is noteworthy here is that the actions denoted by the verbs filling the gerundive slot are typically quite specific in nature, referring to some basic types of events e.g. 'writing', 'talking', 'fishing', 'walking', 'knitting'. In other words, these are the events which are easily paired with specific, concrete images or concepts. This point should become more transparent if we compare these actions with some agentive uses of lexemes occurring in the infinitival slot. Let us, for instance, consider *support* which ranks relatively high among the collexemes of CONTINUE_TO. Here are some examples:

(1) Food shortages now loom in the horn of Africa again, and we will **continue to support** emergency operations.
(2) We will **continue to support** the work of the voluntary sector and promote volunteering.
(3) We **continue to support** activities which will improve the quality of life of poor people

Although the events evoked by *support* in the examples above are all actions, they differ from most of the events denoted by the lexemes listed in Table 6. The point is that *suppor*t construes the events it denotes with much greater degree of generality or abstraction. In other words, the lexeme constitutes what can be referred to as a conceptual shell for a range of different actions, the precise nature of which remains largely underspecified. There are some other collexemes of the infinitival construction which seem to construe the events they denote in a similar way, that is as abstract, generic categories i.e. 'dominate', 'maintain', 'apply', 'monitor'. It should be emphasized here that the distinction between the abstract (or generic) and specific (or more concrete) construals is not to be seen as all-or-nothing but rather as a matter of degree. Hence, the general pattern as observed in the corpus data is also a matter of degree: the greater the degree of schematicity with which an even is construed, the more likely is the use of the infinitival complementizer. By the same token, the greater the degree of specificity or precision with which an event is construed, the more likely is the occurrence of the gerundive complementizer.

Yet another point of interest is the occurrence of *living* among the top collexemes of the gerundive CONTINUE construction. The fact that a lexeme

denoting a stative situation ranks as one of the most significant collexemes of the gerundive complementizer seems at odds with our main observation that stative events are much more likely to co-occur with the infinitival constructions. Consequently, a more fine-grained analysis is in order here. A manual inspection of the relevant concordance reveals that in the vast majority of cases, that is in 23 out of the total of 26 occurrences of the construction under consideration, the matrix verb *continue* takes the infinitive form e.g.

(4) Previously those living with their parents would have been expected **to continue** living there.
(5) Many are able **to continue** living as they wish in their own homes.
(6) The couple plan **to continue** living in Darlington after the shop closes.

The pattern in question suggests that the gerundive form is used here for stylistic reasons, that is to avoid the co-occurrence of two infinitival constructions (i.e. to continue to do sth). Consequently, the semantic factor seems to be irrelevant at this point and the collexeme 'living' should not be seen as a semantic determinant of the gerundive CONTINUE construction.

4.5 CEASE_TO vs. CEASE_ING

Finally, let us consider the collexemes of the two CEASE constructions.

collexeme	freq. in cease_to	freq. in cease_ing	Fisher Yeates
be	696 (631)	13 (78)	5,13E-030
exist	191(170)	0 (21)	2,69E-011
have	75 (67)	0 (8)	1,00E-004
amaze	28 (25)	0 (3)	3,38E-002
function	35 (32)	1 (4)	7,23E-002
apply	23 (21)	1(3)	2,26E-001
hold	11 (10)	0 (1)	2,66E-001
play	9 (8)	0 (1)	3,39E-001
think	8 (7)	0 (1)	3,82E-001
matter	8 (7)	0 (1)	3,82E-001
care	7 (6)	0 (1)	4,31E-001
become	7 (6)	0 (1)	4,31E-001
enjoy	6 (5)	0 (1)	4,86E-001
serve	6 (5)	0 (1)	4,86E-001
occupy	5 (4)	0 (1)	5,48E-001

Table 7. Top 15 collexemes distinctive for CEASE_TO

collexeme	Freq. in cease_to	Freq. in cease_ing	Fisher Yeates
trading	9	50	6,76E-041
feeding	1	5	9,65E-005
using	11	9	1,40E-004
fighting	1	4	7,26E-004
writing	0	2	1,27E-002
brewing	0	2	1,27E-002
firing	0	2	1,27E-002
smoking	0	2	1,27E-002
ringing	0	2	1,27E-002
defending	0	2	1,27E-002
swinging	0	2	1,27E-002
attacking	0	2	1,27E-002
making	6	4	1,93E-002
providing	3	3	2,20E-002
operating	12	5	3,50E-002

Table 8. Top 15 collexemes distinctive for CEASE_ING

The data to be found in Tables 7 and 8 clearly corroborate what has been established for all the other constructions discussed in this paper. As can be seen, there is a very distinct analogy between the infinitival complementizers of CEASE and CONTINUE in that the top three collexemes are exactly the same for both forms i.e. 'be', 'have', 'exist' (with the exception that in the case of CEASE 'exist' comes before 'have' whereas in the case of CONTINUE it is the other way around). It is also worth pointing out that all the other collexemes are much less significant in terms of their strength of attraction to the infinitival slot. Consequently, it can be hypothesized here that the main function of CEASE_TO construction is to encode the cessation of stative, stable situations (or imperfective processes). Among these less significant collexemes can be found items denoting emotional/cognitive states (i.e. 'amaze', 'think', 'care', 'enjoy', 'matter') and a transitional verb 'become'. There are also a few items that seem to evoke agentive events i.e. 'apply', 'hold', 'play', 'serve', 'occupy'. However, as in the case of the CONTINUE constructions, the items in question should be classified as 'conceptual shells', which construe the events they encode as abstract, schematic categories. To illustrate the point, let us take a look at the verb 'play'. Consideration of the relevant concordance, reveals that it occurs almost exclusively in a fixed collocation 'play a role' e.g.

(7) Two years later, the establishments **ceased to play** a major managerial role.

'Playing a role' acts here as a generic category for different actions the precise nature of which is, however, outside of the scope of the predication.

If we now take a look at Table 8, it should become clear that all the collexemes of the gerundive construction denote agentive events and that these events are construed at a level of much greater specificity or precision than those encoded by the infinitival form. Additionally, what should be noted here is that there is a very substantial difference in the significance values between the most strongly attracted collexeme 'trading' and all the other ones, which are only marginally significant. This most significant collexeme construes the situation it denotes in terms of a dynamic succession of distinct acts of buying and selling (typically stocks and shares). Consequently, what is foregrounded here is the dynamicity and iteration associated with this type of agentive behavior.

Given the differences is significance between particular collexemes as observed for the two CEASE constructions, the semantic contrast between the two most significant items to be found in Tables 7 and 8, that is 'be' and 'trading', can be taken as an important determinant of the prototypical semantic function of these two complementizers.

5. Discussion and Conclusions

The research presented in this paper has been founded on the assumption that one can study the semantic properties of abstract syntactic constructions on the basis of their significant collexemes, that is the lexemes which occur in those constructions more frequently than expected. Indeed, the distributional data presented in this paper show that such an analysis provides essential information concerning the nature of the semantic contrast between the two non-finite English complementizers: the gerund and the to-infinitive.[vii]

The distinctive collexeme analysis of the aspectual constructions outlined in the sections above shows that the infinitival and gerundive constructions with the same matrix verb do not represent arbitrary alternations but are semantically motivated. As has been seen, the constructions show a striking consistency with respect to the types of situations encoded by their complements. In most general terms, the distinction between the infinitival and the gerundive complementizer is realized as, respectively, the distinction between stative situations (including cognitive/ emotional states) and non-agentive processes of change/ transition, on the one hand, and agentive events, on the other.

Taking a more detailed look at the infinitival construction, the role of the introductory marker *to* constituting an integral part of the construction cannot go unnoticed. The origins of this marker (e.g. Traugott 1992), should be traced back to the prepositional *to*, and more specifically to its spatial sense *towards*, which in the course of time has been extended to include its use as a preposition of purpose. As has been argued in cognitive linguistics literature (e.g. Langacker 1991; Duffley 2006) prepositional uses of *to* have their roots in the PATH imagery, which has the following schematic structure:

Figure 3. PATH image schema

PATH image-schema as presented above consists of three elements: the source location A, the target location B (a goal) and a vector connecting these two, which at the most prototypical level corresponds to the path traversed by a given entity from A to B. PATH imagery also seems to provide the basic structure for the use of the infinitive marker *to* by means of the conceptual analogy that exists between moving from location A to B and a transition from one state of affairs into another state of affairs. This hypothesis finds its expression and confirmation in the observed, steady tendency of the infinitival construction to co-occur with items denoting stative situations or change of state events. Hence, the infinitival complementizer as used in the aspectual constructions evokes what can be referred to as a *transitional construal*, which motivates the distribution of the BEGIN_TO, START_TO, and CEASE_TO construction. This type of construal corresponds to moving from location A to B. The CEASE_TO construction, which encodes perpetuation through time of a stable situation, at the most schematic level corresponds to the vector connecting the source and the target location of the PATH image-schema.

The gerundive complementizer, on the other hand, co-occurs with dynamic, agentive events and shows a marked tendency to profile *iterative construal* (i.e. series of events). However, although this tendency is strongly marked in the data discussed in this paper, it is not universal. A preliminary manual analysis of particular concordance lines reveals that there are collexemes which encode single or one-off events, that is actions taking place once, at one time and one place.[viii] For instance:

(8) He reached for a sheet of paper, picked up a pen and **began writing.**

Hence, iteration should not be considered as a semantic invariant of the gerundive construction but rather as its prototypical characteristic.

At a more general level, the distinction between the two complementizers discussed in this paper seems to be compatible with the basic distinction made by Langacker (2002: 85–88) between imperfective and perfective processes, which are defined by analogy to uncountable and countable nouns respectively. More specifically, on Langacker's account an imperfective process is internally homogenous (its component states are identical) and unbounded (indefinitely expansible/ contractible). A perfective process on the other hand, is heterogeneous, bounded in time within the scope of predication and divisible or replicable by analogy to individuated things. Consequently repetitive aspect is possible for perfective processes. As specified by Langacker, his imperfectives class is equivalent to Vendler's states whereas perfectives constitute a broad category embracing Vendlerian activities, accomplishments and achievements. Even more important in Langacker's

characterization of process types is the notion of change. As argued by the author:

> (…) a perfective process portrays a situation as changing through time, while an imperfective process describes the extension through time of a stable situation […]. Canonical perfectives (e.g. jump, kick, learn, explode, arrive, cook, ask) clearly involve some change through time. By contrast imperfectives, (e.g. resemble, have, know, want, like) are plausibly interpreted as describing the perpetuation through time of a static configuration.[ix] (Langacker 2002: 86)

Given this characterization, we can conclude that at the most schematic level the distinction between the infinitival and gerundive constructions is construed by analogy to Langackerian distinction between imperfective and perfective processes. It should be, however, emphasized here that the borderline between these two types of construal is in itself a matter of construal, which is to say that it is not fixed or definite but subject to conceptual shifts or accommodations in the minds of speakers. Nevertheless, the general tendency as captured by the quantitative data remains: if a given situation is conceived of in terms of a transition from one state of affairs into another state of affairs or as a static perpetuation through time, it tends to be coded by the infinitive. If, on the other hand, a situation is conceived of in terms of a succession of dynamic, agentive events (iterative construal), or a single perfective event, the gerund takes over as a coding device.[x]

This perfectivizing function of the gerundive complementizer shares some properties with the process of conceptual *reification*, which pervades human thought and perception. The important connection between these two processes is that they both involve imposing bounded construal on the conceived scene.[xi] In addition, reification tends to be associated with specific and concrete representations (at least much more specific and concrete than abstract entities). This seems to be reflected in the observed tendency of the gerundive complementizer to co-occur with items that denote actions characterized by a high degree of specificity or concreteness. These canonical actions such as 'writing', 'talking', 'running' etc., which evoke concrete images contrast with actions characterized by a greater degree of abstractness, that is those with much fuzzier internal structure i.e. 'support', 'apply', 'dominate', 'function'. As the results of the distinctive collexeme analysis indicate, the lexemes belonging to the latter category are much more likely to co-occur with the infinitival complementizer. This seems to be associated with the level of homogeneity in that actions which are more abstract (that is less specific in terms of their internal structure) are also construed of as more homogenous and hence more imperfective like than more concrete actions with more clearly delineated structure.

Finally, a note should be made here concerning the general quantitative tendencies as presented in Figure 1. An important distinction to be made at this point is between unmarked and marked construals, which are associated with the degree of compatibility that exists between the lexical semantics of the matrix verbs and the schematic semantics of the complementizers. It seems that the verbs such as BEGIN and CEASE are most naturally associated with a transitional meaning (a transition from one state of affairs into another

state of affairs) and as such they are highly compatible with the transitional semantics of the infinitival complementizer resulting from the underlying PATH imagery. By the same token, CONTINUE is very likely to construe the conceived scene as an extension through time of a stable situation ('keep existing') and consequently it is highly compatible with the part of the PATH image-schema which is represented by the vector connecting the source and target destinations, with the source and target locations de-focused (cf. fig.3). Of course, each of the verbs in question has a potential for perfectivizing construal, in which case they denote the onset, continuation or cessation of (iterative) agentive events. However, the lesser degree of compatibility between the lexical semantics of these verbs, on the one hand and the reifying (bounding) function of the gerundive complementizer, on the other is reflected in the lower frequency of co-occurrence of these two elements. The only exception to this tendency are the START constructions, which do not show such a strong association with the infinitival form. This in turn, must result from some unique semantic properties of the verb START, which distinguish it from the nearly-synonymous BEGIN. A detailed analysis of some subtle nuances of meaning between these two verbs should shed some light on the observed quantitative differences in their complementation patterns. This, however, will be left as a subject for future research.

Notes

[i] The research presented in this paper was carried out with the financial support of the Polish Ministry of Science and High Education

[ii] Minimal pairs with the common matrix verb seem particularly useful in accounting for the schematic (abstract) semantics of complement categories. It is due to the fact that the semantics of the matrix verb (being the same for both constructions) does not blur the picture, that is it does not interfere with the semantics of the syntactic categories it co-occurs with, and thereby, the true nature of the semantic contrast between the two constructions becomes more apparent.

[iii] The method has its theoretical underpinning in construction-based approach to language i.e. it treats language as an inventory of constructions defined as form-meaning pairings.

[iv] As argued by Stefanowitsch and Gries Fisher Exact test avoids the shortcomings associated with some other test of collocational significance which could be potentially applied in distinctive collexeme analysis (e.g. T-test, Mutual Information test). The main argument is that Fisher Exact test (being exact) 'does not make any distributional assumptions and therefore does not require any particular sample size' (2004:101). This, in turn, makes it an appropriate statistics for investigating even very sparse linguistic datasets, which is often the case in distinctive collexeme analysis.

[v] Sketch Engine is a corpus query system available from Lexcom Lexical Computing (http://www.sketchengine.co.uk/)

[vi] The correlation is such that the smaller the value of the Fisher test the greater the strength of attraction between a construction and the lexemes that fill it.

vii The reservation to be made here is that the research presented in this paper has been limited to some broad semantic distinctions based on the lists of collexemes yielded by the collostructional analysis. The point however is that a full and comprehensive analysis should take into consideration the polysemous nature of particular collexemes. Consequently, a more fine-grained analysis aimed at disambiguating the polysemous lexical items is in order. A detailed study of this kind however goes beyond the scope of the present paper. For a preliminary attempt, see Kaleta (2011, available from the author).

viii The exact proportion of such uses needs to be established in a detailed, fine-grained analysis of every single collexeme, but this is a task that goes beyond the scope of this paper.

ix The notion of change as used by Langacker with respect to perfective processes should be distinguished from the way it is used in this paper to talk about change of state situations. On Langacker's account 'change' constitutes an inherent part of the semantic value of lexemes referred to as perfectives. In my account the notion of change is restricted to the process of transition between two situations viewed as stative (or imperfective) and not as an inherent property of those situations. Consequently, there is no contradiction between Langacker's conception of imperfectives as stable situations and my use of this term with respect to the infinitival complementizer and its stative collexemes.

x The importance of quantitative data cannot be overestimated in the context of the present discussion. More specifically, given that meaning is a matter of construal, patterns that prevail in linguistic behavior are an important indication of prototypical semantic representations, which should be distinguished from more peripheral ones. Quantitative data are also important in the context of Stefanowitsch and Gries's (2003:213) observation that semantic compatibility of schematic syntactic constructions and their lexical fillers does not necessarily mean semantic identity. As follows from the case studies carried out by the authors, the collexemes that are most strongly attracted by particular constructions are at the same time most representative for those constructions. The less significant collexemes, on the other hand, can be attracted by a given construction by the process of prototype extension or family resemblance and as such are not necessarily a proper basis for generalizing about the semantic content of particular constructions.

xi The notion of bounding calls for some specification here. As has been argued after Langacker, boundedness is a property of perfective processes. Also conceptual reification seems to share this property. However, there are authors who believe that reified things are construed of in terms of mass nouns (e.g. Taylor 2002: 399) and as such are devoid of any boundaries. The point, however, is that mass nouns prototypically denote substances or liquids which do not exist or float unrestricted but come in different kinds of containers (either human-made or natural e.g. lakes surrounded by shorelines or seas whose spatial extension is visually restricted by the horizon). Although this boundary is not an inherent property of the entities denoted by mass nouns and as such is not a part of their profile, it is still there in the background. Consequently, it should be concluded here that reification or nominalization always involves imposing bounded construal – the boundaries being either a part of a concept's profile (in the case of countable nouns) or de-focused (implicit) but still present in the scope of predication (as is the case with mass nouns).

References

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

Dirven, R. (1989). A cognitive perspective on complementation. In D. Jaspers et al. (eds.), *Sentential Complementation and the Lexicon: Studies in Honour of Wim de Geest*. Dordrecht – Holland/Providence – U.S.A.: Foris Publication. pp. 113-139.

Duffley, P.J. (2006). *The English Gerund-Participle: A Comparison with the Infinitive*. Berkeley: Peter Lang Publishing.

Egan, T. (2008_. *Non-finite Complementation: A usage-based Study of Infinitive and -ing Clauses in English*. Amsterdam/New York: Rodopi.

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

Gries, S. Th. and A. Stefanowitsch (2010). Cluster analysis and the identification of collexeme classes. In J. Newman and S. Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford, CA: CSLI. pp. 73–90.

Kaleta, A. (2011). English non-finite complementizers in aspectual constructions: A distinctive collexeme analysis. Unpublished manuscript.

Langacker, R. (1991). *Foundations of Cognitive Grammar*, *volume II: Descriptive Applications*. Stanford: Stanford University Press.

Langacker, R. (2002). *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. 2nd edition. Berlin/New York. Mouton de Gruyter.

Pedersen, T. (1996). Fishing for exactness. *Proceedings of the SCSUG 96 in Austin*. pp. 188–200.

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

Taylor, J.R. (2002). *Cognitive Grammar* . Oxford: Oxford University Press

Traugott, E. (1992). Syntax. In R.M. Hogg (ed.), *The Cambridge History of the English Language*, *Vol. I*: *The Beginnings to 1066.* Cambridge: Cambridge University Press. pp. 168–289.

Wierzbicka, A. (1988). *The Semantics of Grammar*. Amsterdam/Philadelphia: John Benjamins