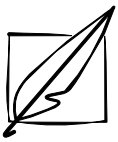


Jan Kochanowski University Press

This is a contribution from *Token*: *A Journal of English Linguistics*
Volume 1/2012.
Edited by Sylwester Łodej and John G. Newman.

On the complementation of *start, begin* and *continue* in spoken academic American English

Oskar Gawlik
Jan Kochanowski University

ABSTRACT

The main goal of this article is to present the results of a linguistic inquiry into the syntax of *start*, *begin* and *continue* in contemporary spoken academic American English. The main goal of the study was to determine the frequency of occurrence of these verbs and their preferred non-finite complements – the gerund form and the present infinitive form – in that variety of English. The incidence of *start, begin* and *continue* was investigated in both small and large lectures in the MICASE corpus. The precepts of corpus linguistics have together served as the major methodological tool. The audience targeted here includes theoretical and applied linguists interested in English linguistics as well as students of the discipline and scholars of related fields.

1. Introduction

There are several good reasons for investigating the preferred complements of *start*, *begin* and *continue* (i.e. the gerund and the present infinitive) in contemporary academic spoken American English. Generally, variation is natural to language, and genre-induced variation in English deserves more attention than it has so far received. More specifically, genre-induced variation in the syntax of the lemmas of *start, begin* and *continue* in spoken academic English, due to genre-specific lexicogrammatical patternings, is one particularly interesting yet apparently neglected area. Most English grammar handbooks lack substantive information concerning the preferred complements of *start, begin* or *continue.* Moreover, such references say little or nothing on genre-specific preferences for the non-finite complements of these verbs (i.e. the gerund and the full infinitive), the exception being

the *Longman Grammar of Spoken and Written English* (Biber et al. 1999). In the absence of descriptive detail, we might assume that these verbal complements are in free distribution irrespective of genre, a notion which has for present purposes been assumed not to be the case. Consequently, striving for naturalness of expression, for example, when delivering an academic paper, and unable to answer my own or my students' questions concerning the most frequent, ergo most natural, verbal complements of *start, begin* and *continue* in spoken academic English, I embarked on a study, a corpus study, into the syntax of these verbs, an enterprise which I, and others, such as Swales (1990), the doyen of English for Academic Purposes (EAP), have resorted to since English grammars and other sources fail to provide unequivocal answers. Furthermore, an attempt has been made to investigate spoken academic English, instead of written academic English, on the grounds that the latter has been explored relatively thoroughly, while the former still appears to be *terra incognita*.

2. Corpus linguistics as a methodological tool

Although corpora were drawn on to investigate language even in the heyday of Chomsky's generative-transformational grammar, corpus linguistics (CL) offers a relatively new methodological tool. It enables us large scale investigation of actual linguistic production, rather than investigation of the results of linguistic introspection, which has not infrequently proven unreliable. CL appears to represent a methodological shift from a focus on competence to one on performance, to use Chomsky's terms, or from one on *langue* to one on *parole*, to employ de Saussure's terms. However, the swing of the pendulum has not applied across the board: a number of scholars advocate complementing corpus-based and corpus-driven methodologies with introspection, particularly in qualitative analyses (e.g. Rusiecki 2006), an approach I also adopt in my linguistic inquiries.

The value of corpus linguistics and its empirical basis is difficult to overestimate, and the wide panoply of linguistic pursuits that it involves results in numerous implications and applications of findings, particularly because of the burgeoning number and variety of corpora being made available. However, in the interests of space, I will not attempt a comprehensive treatment of the merits of CL here, but rather will simply laud one major work which represents the methods and fruits of CL: the *Longman Grammar of Spoken and Written English* (LGSWE; Biber et al. 1999).

This grammar is devoted to an analysis of lexicogrammatical patternings in four genres of current English: academic, journalistic, fictional, and conversational. Its quantitative and qualitative analysis of the characteristics of these four English genres paints a comprehensive picture of English grammar, highlighting the differences between the genres and validating previous (hypo)theses concerning language while disconfirming others, which, understandably, has a number of implications and applications, pedagogical ones among them. Consequently, LGSWE appears to have surpassed other books devoted to the grammar of English inasmuch as it both epitomizes the grammatical theory of CL and (re)validates CL's empirical approach, this last by reference to actual language use, which reveals the employment of certain linguistic structures while eschewing others, genre by genre. No doubt interest in employing CL as a methodological tool is gaining popularity. Tens of English language corpora are being utilized by scholars worldwide today, and increasing numbers of corpus-based or corpus-driven publications are appearing (cf. e.g. Mauranen 2001; Swales 2001; Łyda 2007; Gawlik 2011).

3. Selected studies of academic English

As concerns academic English, a large number of studies have been done in applied linguistics, with practical applications constituting the main rationale for discovering the intricacies of academic discourse. Consequently, linguistic interests in academic English have mostly revolved around the study of written English grammar and lexis, such as those of 'the research article', the academic written genre *par excellence*, with a view to exploring a wide range of linguistic aspects: contrastive rhetoric (Galtung 1981; Clyne 1987; Scollon – Scollon 1995; Connor 1996; Mauranen 1997), genre analysis (Swales 1990), metadiscourse (Hyland – Tse 2004), hedging (Hyland 1998; Varttala 2003) or evaluation (Stotesbury 2003). The results of such linguistic inquiries are often extended to practical applications: 'the research article' is a genre which millions of people, both students and academics, are exposed to and grapple with in the milieu of academia. In sum, explorations into academic written discourse, frequently propelled by a view to practical applications, appear to have borne fruit in the form of numerous works targeted at both experts and student neophytes.

Of late, however, investigations of academic English have bifurcated into analyses not only of the written mode of academic discourse but

also those of the equivalent spoken mode. The latter have stemmed from both a growing cognizance of the need to expand such studies and an increased availability of corpora of spoken academic English, such as the *Michigan Corpus of Academic Spoken English* (MICASE; Simpson et al. 2002). Consequently, research into EAP, particularly into the prototypical genre of spoken academic English, or 'the lecture talk', as exemplified by studies of metatext (Swales 2001), reflexive academic talk (Mauranen 2001), evaluation (Swales 2004), concession (Łyda 2007), and *verba dicendi* (Gawlik 2011), has gained momentum. Still, despite a growing body of research, investigation of the spoken variety of academic English remains in its infancy, which speaks to importance of the present study.

4. Methods of analysis of the occurrence of the non-finite verbal complements of *start*, *begin* and *continue*

The focus in this investigation was on the nearly synonymous verbs *start* and *begin* as well as the verb *continue*, and the patterns of verbal complementation regarding the apparently competing gerundial and infinitival non-finite complements, with a view to determining which of the two is more preferable in spoken academic English. The choice of these two types of complements precluded analysis of any other types of complements, such as the progressive or perfect infinitives. That decision was made for at least two reasons: firstly, the gerund and the infinitive mark a different aspect, and thus convey different nuances of meaning, which implies that they should not be considered complements that are competing with each other in relation to the three aspectualizers analysed. The disregard for other types of complementation also stemmed from the supposition that the results of the study could have specific practical, perhaps pedagogical, applications concerning relevant verbal complement choices.

The synchronic corpus investigation was conducted on the prototypical genre of spoken academic English, the lecture talk, and relied on data from MICASE, a collection of authentic and unscripted texts which were recorded at the University of Michigan and which consequently represent the American variety of English. The corpus consists of almost 200 hours (approximately 1.8 million words) of contemporary academic speech, divided into sixteen different speech events, encompassing topic-matter of both the humanities and the sciences. The corpus can be accessed using a wide variety of sociolinguistic filters. For instance, one may search

for data in speech events of women only or men only. Unfortunately, for all the merits of the corpus, it is not tagged for grammatical categories, which meant that the data culled needed to be checked manually for categorical sorting in order to avoid arriving at skewed results.

The MICASE data was complemented by data of the *Corpus of Contemporary American English* (henceforth COCA) complied by Mark Davies at Brigham Young University in 2008. That corpus is the largest contemporary body of American English, and it spans the years 1990–2011. COCA is annotated biannually, and currently it consists of approximately 425 million words of five genres: spoken, fictional, journalistic-magazine, journalistic-newspaper, academic. Grammatically tagged, dialectally representative, inclusive of spoken academic data, comparatively large, the corpus was invaluable to the present study.

5. Previous inquiry into complementation

By now, a number of studies on complementation, both synchronic and diachronic, have employed the methodology of CL (cf. e.g. Rudanko 2000; Egan 2008; Mair 2009). However, before CL began to develop in earnest, statements about complementation which were based on introspection and few instantiations were, to an extent understandably, the norm. One such statement regarding two of the aspectualizers examined, *start* and *begin*, is this:

> An informal survey of native speakers of English indicates that most believe *begin* and *start* to be close synonyms and almost entirely interchangeable. Some say only that *begin* seems slightly more 'formal'. This feeling may be due to the fact […] that *begin* occurs in a more restricted number of contexts than *start*. A careful analysis of these two aspectualizers, however, turns up a surprising number of differences […]. [T]here must exist semantic as well as syntactic distinctions between these aspectualizers which native speakers attest to by their unselfconscious and natural use of them [sic.]" (Freed 1979: 68).

The claim here that the two verbs are virtually interchangeability is typical of traditional grammar handbooks and usage guides. More interesting is the hypothesis that the number of contexts of the respective nouns differs, and the suspicion that semantic and syntactic restrictions must also affect the relevant

behavior of these verbs (and allow them to avoid obsolescence). It is from the point of such suppositions that we may proceed by the means of CL and attempt to establish patterns of restriction and preference, patterns among instances of non-finite verbal complementation specific to these verbs (and, for present purposes, *continue* as well), with a greater measure of certainty.

6. The syntax of *start*, *begin* and *continue* in light of empirical investigations

The synchronic corpus investigation of the gerundial and infinitival complementation of the lemmas of *start*, *begin* and *continue* in the MICASE, yielded the results included in Appendix 1. The pertinent data there, after lemmatization, include all and only verbal occurrences of the word forms, and all and only constructions containing non-finite gerundial or infinitival complements. Because the MICASE set of lectures, small and large, consists of 584,970 words, the figures in Appendix 1 are normalized to one million occurrences for ease of interpretation. For ease of comparison, the results are also presented in Appendix 2 in a graphic form.

Statistically, *start* is the most frequent of the verbs, and it is most often followed by a gerundial complement, a total of 171 times (approximately 292 occurrences per million words). By comparison, infinitival complements of *start* are less numerous, totaling 120 instances (approximately 205 occurrences per million words). Items (1) and (2) below exemplify the two respective types.

(1) you had to memorize? okay, you have a lot of A-T-P molecules when you're living on glucose. if you **start** depleting the glucose you're gonna be running out of A-T-P and available A-T-P will get converted in. (LES175SU079)

(2) assigned to that descriptor. um, i think, here experience comes to bear. i think after a while you **start** to see certain descriptors, become familiar with them, see them more and more, and find, ah this will. (LES335JG065)

Begin, by contrast, exhibits a weak preference for the gerund, with only 7 instances (12 occurrences per million words), and a strong preference for the infinitive, with as many as 84 instances (143 occurrences per million words). Examples of those instances are seen here:

(3) thousands can vote, and people are rigging votes in all kinds of exciting and interesting ways, this **begins** to matter. there is a feeling that, you never get to vote. your vote never counts. what ma- who care. (LEL215SU150)

(4) and federal governments, personnel officers, they **begin** entering graduate programs. they'd already **begun** entering medical schools already eighteen fifties uh som- women and law schools becomes i. (LEL105SU113)

Continue, like *begin*, is more frequently complemented by the present full infinitive than it is by the gerund. The figures, respectively, are these: 46 occurrences (79 per million words) versus 4 occurrences (7 per million words). Examples of each follow.

(5) than a hundred miles of country before reaching the sea, into which they plunge, unhesitatingly, and **continue** to swim on until they die. even then they float so that their dead bodies form drifts, on the seasho. (LEL175SU112)

(6) right but why can't you just say sit still, and, **continue** making money (xx) cuz somebody. (LEL 565SU064)

As these results indicate, when set beside *begin, start* is the more frequent word (near-synonym) in spoken academic American English. Not distantly related to the respective frequencies of these two verbs is a pattern regarding comparative formality. The fact that both words occur in this corpus of formal English, and manifestly in not wildly different quantities of instances, argues against a significant difference in usage in terms of formality, and against the supposition by Freed (1979: 68) quoted above. Of course the time of a generation has passed since that supposition was stated. The results also suggest that *start* is typically complemented by the gerund form, while *begin* is typically complemented by the full infinitive, with the gerund form appearing to constitute an exception.

My corpus findings tally with those of other reseachers with respect to *begin* as "[one of] the most common verbs controlling to-clauses", and *start* as "[one of] the most common verbs controlling *ing*-clauses" (Biber et al. 1999: 699). However, as regards the complements of *continue*, my findings stand in contrast. Biber et al. (1999: 741) claim that "[a]apectual verbs are common

with *ing*-clauses, because their meanings concern the delimitation of actions that go on over time – relating to the starting point (e.g. *start*, *postpone*), the end point (e.g. *stop*, *quit*), or their progress (e.g. *keep on*, *continue*, *resume*)". This general claim does not appear to be applicable to spoken academic American English, as my corpus findings testify to the contrary: *continue* is typically complemented by the present to-infinitive in the MICASE lectures, which, by implication, might constitute a style marker of spoken academic English.

Although my original intention was to investigate four, rather than three, aspectual verbs, the fourth being the semantically proximal *commence*, on searching for tokens of *commence* in the MICASE, it emerged that there was not a single occurrence of any form of this verb either in the lecture talk analysed or in any other speech event contained in the corpus, which, given the verb's comparatively formal connotation, was unexpected. Since this fact ran counter to my introspective judgement, I sought evidence of *commence* and its complementation in a much larger corpus of American English, i.e. the *Corpus of Contemporary American English* (COCA), which currently consists of approximately 425 million words, in order to amass some genre-varying data on the verb and to confirm or repudiate the relatively low frequency of the verb evident in the formal MICASE data. Appendix 3 illustrated the COCA data culled.

There are at least two obvious conclusions to be drawn from the results included in Appendix 3. First, *commence* is of comparatively low general frequency, ranging from 1.3 to 6.96 tokens per one million words. Second, *commence* was more frequently employed in written academic English than it was in any of the other four genres, scoring 597 hits here, which translates into its 6.96 tokens per one million words. Of course, this implies little or nothing about its relative popularity in spoken academic American English. However, the tendency might suggest an overcoming of field and tenor, to frame the statement in Hallidayan parlance. Although the results need validation, what the corpus findings appear to have corroborated is that *commence* is infrequent not only in spoken academic English (as attested by MICASE), but also in certain other genres, as confirmed by the findings obtained from COCA. Although we might reasonably expect to encounter the comparatively formal verb *commence* frequently in formal spoken academic English, followed there by *begin* and *start*, in descending order of incidence, the evidence gleaned indicates that spoken academic American English discourse prefers *start* before *begin*, and both of these aspectualizers before

that of *commence*, and by extension, then, suggests that spoken academic discourse lies, in terms of formality, between written academic discourse and everyday spoken discourse, a pattern identified elsewhere (Swales 2004).

7. Conclusions

In light of the evidence presented above, it may be concluded on grounds of their respective frequencies of occurrence and patterns of complementation that the two aspectualizers *start* and *begin* are not absolute, but rather proximate, synonyms of spoken academic American English. In terms of frequency, *start*, with 291 hits in MICASE, is clearly a more popular option than *begin*, with its 91 hits. As regards syntax, *start* selects gerundial complements (171 hits in MICASE) more commonly than it does infinitival ones (120 hits), and *begin* manifests the opposite pattern, selecting the full infinitive (84 hits) far more often than the gerund (7 hits). The verb *continue* also prefers the infinitival complement (46 hits) to the gerundial (4 hits) in MICASE (see Fig. 1).

These differences in frequency and syntax may imply contrasting style marking by *start* and *begin* in spoken academic English. Investigation of the overall incidence of these two verbs, conducted from semantic and pragmatic viewpoints, would certainly further our understanding of their behavior in more formal as well as less formal contexts.

REFERENCES

Sources

Corpus of Contemporary American English (COCA) Compiled by Mark Davis (Brigham Young University). http://corpus.byu.edu/coca/.

Simpson, R.C. et al.
2002 *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Special studies

Biber, Douglas et al.
1999 *Longman Grammar of Spoken and Written English*. Essex: Longman.

Clyne, Michael
1987 "Cultural differences in the organization of academic texts", *Journal of Pragmatics* 11, 211-247.
Connor, Ulla
1996 *Contrastive Rhetoric. Cross-Cultural Aspects of Second-Language Writing.* Cambridge: Cambridge University Press.
Egan, Thomas
2008 *Non-Finite Complementation: A Usage-Based Study of Infinitive and -ing Clauses in English*. Amsterdam: Editions Rodopi.
Freed, Alice F.
1979 *The Semantics of English Aspectual Complementation.* Dordrecht: Reidel Publishing Company.
Galtung, Johan
1981 *Structure, Culture and Intellectual Style: An Essay Comparing Saxonic, Teutonic, Gallic and Nipponic Approaches*. London: SAGE.
Gawlik, Oskar
2011 *Basic Verba Dicendi in Academic Spoken English. The Syntax, Semantics,* Pragmatics and Functions of Communication Verbs: Say, Talk, Tell and *Speak.* Saarbrucken: Lambert Academic Publishing.
Hyland, Ken
1998 *Hedging in Scientific Research Articles.* Amsterdam: John Benjamins.
Hyland, Ken – Po Ting Tse
2004 "Metadiscourse in academic writing: A reappraisal", *Applied Linguistics* 25(2), 156-177.
Łyda, Andrzej
2007 *Concessive Relation in Spoken Discourse*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
Mair, Christian
2009 *Infinitival Complement Clauses in English. A Study of Syntax in Discourse.* Cambridge: Cambridge University Press.
Mauranen, Anna
1997 *Cultural Differences in Academic Rhetoric.* Frankfurt: Peter Lang.
2001 "Reflexive academic talk: Observations from MICASE". In R.C. Simpson et al. (eds.) *Corpus Linguistics in North America. Selections from the 1999 Symposium.* Ann Arbor: University of Michigan Press, 165-178.
Rudanko, Martti Juhani
2000 *Corpora and Complementation: Tracing Sentential Complementation Patterns of Nouns, Adjectives and Verbs over the Last Three Centuries.* Lanham, Maryland: University Press of America.
Rusiecki, Jan
2006 "Badania korpusowe a 'lingwistyka gabinetowa': czy te dwa podejścia kiedyśsię spotkają?". In A. Duszak et al. (eds.). *Korpusy*

w angielsko-polskim językoznawstwie kontrastywnym. Kraków: UNIVERSITAS, 17-29.

Scollon, Ronald – Suzanne Scollon

1995 *Intercultural Communication: A Discourse Approach.* Oxford: Blackwell.

Stotesbury, Hilkka

2003 "Evaluation in research articles abstracts in the narrative and hard sciences", *Journal of English for Academic Purposes* 2, 327-341.

Swales, John

1990 *Genre Analysis. English in Academic and Research Settings.* Cambridge: Cambridge University Press.

2001 "Metatalk in American academic talk. The cases of *point* and *thing*", *Journal of English Linguistics* 29(1), 34-54.

2004 "Evaluations in academic speech: First forays." In G. d. E. Camiciotti and E. T. Bonelli (eds.) *Academic Discourse – New Insights into Evaluation.* Bern: Peter Lang, 31-53.

Varttala, Teppo

2003 *Hedging in Scientifically Oriented Discourse. Exploring Variation According to Discipline and Intended Audience*. Ph.D. dissertation. Tampere: University of Tampere, Finland.

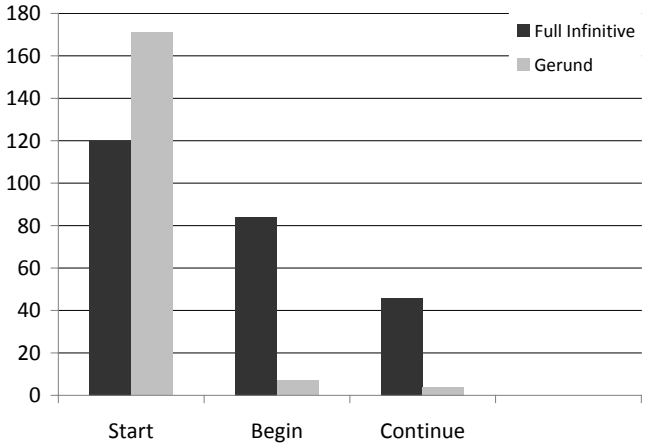
APPENDIX 1

The gerund versus the infinitive as complements of *start*, *begin* and *continue* in the MICASE corpus of lectures

The verb investigated	Total occurrences	Occurrences after lemmatization	Complementation by means of the full infinitive		Complementation by means of the gerund	
			Absolute number of tokens	Number of tokens per one million words	Absolute number of tokens	Number of tokens per one million words
Start	654	291	120	205	171	292
Begin	262	91	84	143	7	12
Continue	104	50	46	79	4	7

APPENDIX 2

The occurrences of the gerund versus the full infinitive as complements of the tokens of *start*, *begin* and *continue* in the MICASE corpus of lectures.



APPENDIX 3

Deployment of tokens of *commence* in different genres in contemporary American English (COCA data).

Genre	Total number of tokens	Frequency of occurrence per one million words
Spoken	117	1.3
Fiction	497	5.85
Magazine	335	3.71
Newspaper	187	2.16
Academic	597	6.96