

## Can What Counts in Complexity Be Counted?

Wayne Rimmer

The motivation for this corpus-informed study is the weak theoretical basis in the literature for evaluating grammatical complexity (GC). Methodologies with a quantitative focus have dominated enquiry into GC, with unit length and subordination being awarded a special status. Written data from the International Corpus of Learner English is used to debunk the notion that quantitative units fully account for GC. Examination of individual texts reveals that GC consists of how language is selected, combined and patterned to give writing purpose and direction. Quantitative methods such as T-units are not sensitive to language as it is used in specific contexts and hence fall short of capturing the delicate mechanisms which engineer GC. A case is made for a sounder notion of GC, an approach marked by a commitment to qualitative analysis.

### 1. Introduction

A key difference between first and second language acquisition is that the former is almost invariably successful while the latter is characterised by very pronounced variances in achievement. It is convenient to view these gradations in second language acquisition as on a cline from zero knowledge to ultimate attainment. One factor often singled out as operating towards the higher end of this scale is grammatical complexity (GC). With its associations of sophistication and abstruseness, GC is seen as a sign of linguistic progress and accomplishment. From an applied linguistic perspective, empirical investigations into GC have been influential in identifying the nature of a construct so closely identified with superior performance. However, existing research has struggled to account for GC in a way that realises its potential for practical applications such as pedagogy. This paper discusses the issues involved in describing GC. It argues that attempts to define GC using quantitative measures have only met with partial success, leaving the nature of the construct obscure, and, drawing on corpus-informed research, suggests a more promising direction to explore the phenomenon.

It is important at the outset to delimit the discussion to that of complexity of form, rather than psycholinguistic complexity. This study is confined to complexity of form; psycholinguistic factors such as the storage and retrieval of grammatical items are largely ignored. The decision to concentrate on form is motivated by the commitment to a corpus methodology. Corpora are not real-time phenomena, they are essentially historical records, and as such they do not lend themselves readily to psycholinguistic research, which typically has an experimental approach, as in Lazarte and Barry's (2008) demonstration of how highly-embedded sentences impede reading comprehension in English and Spanish. This is not to dismiss the significance of psycholinguistics, and certainly not to reject the connection between complexity of form and its representation in the mind. However, the psycholinguistic dimension of GC lies beyond the scope of the present paper.

Similarly, the study is restricted to second language writing. The written focus is again in part due to methodology, writing being much easier to collect and analyse than oral samples. In addition, there is certainly an assumption that speaking is formally less complex than writing, making it less worthy of attention, although this contrast is problematic. For example, while it is often posited (e.g. Carter & McCarthy, 1997; Willis, 2003) that speaking

prefers coordination to subordination as a means of organising clause structure, oral language seems resistant to classification in these terms: Brown (2006) evaluating samples of different proficiency from the IELTS speaking test found that clausal density has no impact on scores. Speaking lends itself less readily to formal analysis, and, as Hunston (2002) suggests, it needs new tools of description. Certainly, the vast majority of the literature on GC, and everything cited hence on, examines writing rather than speaking.

Two features have dominated enquiry into GC: unit length and subordination. In a constituent phrase-structure grammar, the relevant units are at the level of sentence, clause and phrase. Measured in word length, longer units are claimed to represent greater complexity. Subordinate clauses are favoured presumably because hypotactic relationships suggest more ingenuity than paratactic links. Both unit length and subordination are essentially quantitative methods of GC in that they can be counted, compared and subjected to statistical analysis.

Numerous studies have done just that. The following are all corpus-based examples: Aronsson (2003) on pseudo-cleft sentences; Kennedy and Thorp (2007) on sentences; Rimmer (2008) on noun phrases; Sampson (2003) on 15 types of subordinate clause. Such studies, even if not directly concerned with GC, have contributed valuable information to their respective areas of interest, as well as SLA generally, but the underlying assumption, that more is better, is questionable. It has not been demonstrated that GC is an incremental phenomenon which can be counted and calibrated in terms of pure volume. The purpose of this study is to argue that quantitative measures offer a skewed account of GC, one that under-represents the construct unless a qualitative dimension is added to the methodology.

## **2. Quantitative measures of grammatical complexity**

The validity of correlating unit length with GC depends on identifying a stable unit for comparison. This is problematic, best illustrated by the difficulty of even defining a unit even as basic as the sentence: Matthews (1981: 26-49) devotes most of a chapter to the conundrum. The literature is thus replete with a range of complexity measures. To illustrate with a recent and very well-cited secondary study, Ortega (2003) reviews twenty-five studies containing quantitative data rating the syntactical complexity of written work of advanced level second language students. The measures employed, in order of frequency, were mean length of T-unit (MLTU), mean length of clause (MLC), mean length of sentence (MLS), mean number of T-units per sentence (TU/C), i.e. coordination, mean number of clauses per T-unit (C/TU), and a similar measure of subordination, mean number of dependent clauses per clause (DC/C). By aggregating results from the studies, Ortega reports critical magnitudes which are statistically significant for determining between-proficiency differences in syntactical complexity. For example, for MLTU a difference of two or more words per T-unit indicates the subjects are from separate samples.

The attractiveness of the methodology is obvious for by the relatively simple technique of counting words/clauses across samples a numerical cut-off point between populations is obtainable. However, a purely quantitative approach offers only a partial solution to the problem. First, the relevance of the measures is suspect. To take the T-unit (an independent clause plus any dependent clauses) this measure suffers from severe construct interference in that it combines lexical and grammatical competence. If GC is the construct, grammar not vocabulary should be measured. This principle is impossible to maintain when counts are based on words. To illustrate, two T-units are presented.

- (a). He felt great after the finish. (single clause)
- (b). He felt great after he had finished. (main + subordinate clause)

Here (b) has a higher MLTU and in general the degree of subordination would increase MLTU. However, this tendency is not a constant. *The runner felt great after the finish* would contribute to the same MLTU as sentence (b) although it is a single clause. Inserting vocabulary items into a T-unit will dramatically increase MLTU without affecting clause structure.

The delighted runner really felt absolutely great after finishing the grueling two-hour marathon race. (main + subordinate clause)

This sentence is obviously contrived but there is a genuine worry that MLTU measures lexical rather than grammatical knowledge since users with a rich vocabulary but poor range of structures could achieve an artificially high MLTU by stringing together modifiers. The same reservation applies to mean length of clause and sentence because they too are based on word counts.

The second objection, and the most critical, is that the methodology allows no calibration or evaluation of language use. The numerical values reported as critical magnitudes do not reflect the quality of the sample because they conflate the diversity of clause structure into prototypical patterns. Thus, DC/C and C/TU do not distinguish types of subordination and the two sentences below would be of equal weight.

- (a). Why was Amy so stupid that she was expelled?
- (b). Why was Amy so stupid that it's unreal?

Superficially, the sentences are similar: *Wh*-adjunct + interrogative main clause + finite dependent clause. Actually, the status of the sentences should be judged very differently. Syntactically, the subordinate *that*-clause functions as comparative complement in (a) and adverbial in (b). The meaning of (a) is 'What is the reason that Amy didn't avoid expulsion?' and (b) 'Amy's stupidity is puzzling'. Prosodically, there would also be a difference, with (a) comprising a single tone unit and (b) two. In terms of usage, the (a) construction, *so* + adjective + finite clause, is much more frequent than (b). There are few constructions like (b) when a declarative *that*-clause can be used as an adverbial (another example is the somewhat archaic *that*-adverbial of purpose as in the war memorial epitaph *We died, that you may live*). The linguistic significance of (b), whether *that*-adverbials are restricted in register or even marginal, is actually of less interest here than the methodological implications. The type of diversity and low-frequency usage represented by (b) should be accounted for, not ignored. Pinker (2004: 951) makes the point that we need "to *document* the putative rarity of sentence constructions" (emphasis in original). There is no way that purely quantitative methodologies can do this for they aim to establish patterns and similitude rather than identify and evaluate genuinely interesting departures from the norm. There is no external criterion involved, such as frequency or corpus attestation, so they are weak indicators of the value of individual constructions and their relationship to norms.

What is more, statistics can hide or, worse, distort the true nature of performance, as illustrated by a study (Iwashita et al. 2008) of GC as one component of second language speaking proficiency. Iwashita et al used four measures of GC to investigate features of test-taker language at the five attainment levels of the ETS-TOEFL test. The first three are familiar from Ortega (2003): C/TU, DC/C and MLU. The fourth is verb phrases per T-unit. They found that the three measures of clausal density – i.e. C/TU, DC/C and verb phrases per T-unit, but not MLU - correlated very poorly with performance. For example (Ortega 2003), candidates with the minimum Band 1 score actually recorded a higher C/TU value than candidates with the maximum Band 5 score. Taken at face value, this suggests GC is unrelated to proficiency. However, the results are deceptive. Because the figures are reported as ratios (cf. Ortega), they do not reflect the patent increase in the volume of subordinate clauses that accompanies higher performance. More capable candidates generally produce more subordination and more developed T-units, but this tendency is cancelled out by ratios.

In a sample of individual results (Ortega 2003: 45), a Band 1 candidate produced two T-units with a total of three clauses; a Band 5 candidate produced twelve T-units with a total of nineteen clauses. Both ended up with very similar C/TU values of 1.5 and 1.58 respectively. The authors also note the high standard deviation for GC across all five ability bands. Candidates credited with the same level were actually performing very differently in terms of the density of subordination incorporated into their speaking. Clearly, comparison of group scores in this study, especially in the form of ratios, is unhelpful when attempting to differentiate between levels of performance.

In the discussion above, subordination figures heavily as an *ipso facto* marker of GC. This is confirmed by the many references in the literature that equate subordination with GC (e.g. Huddleston 1984; Quirk et al. 1985; Towell 2002; Willis 2003; Purpura 2004; Carter & McCarthy 2006). However, there is some confusion between developmental stages, where subordination is acquired relatively late (see Pienemann 1998) and text production, which may not feature subordination at all yet still be grammatically complex. In other words, the order of acquisition of language items, including subordinate clauses, has no direct bearing on their complexity. To illustrate this, Berman (2008) compares L1 competence on variables of first language, modality (speech/writing), genre (e.g. types of narrative) and age group (primary, secondary and adult). Berman reports a clear pattern of development in the density and function of clauses in writing and speaking. The results broadly support a progression from coordination to subordination patterns in text. The most advanced stage, typically realised only in adolescence, is 'nesting', i.e. the stacking of different clause types together, for example the co-ordination of complement clauses or the complementation of adverbials. This skill is wedded to the awareness of discourse and general cognitive development. Complex sentence construction, or 'clause packaging' to adopt Berman's terminology, thus signals the connection between language and socio-cultural conventions and conditions of use:

Clause packaging reflects linguistic command of complex syntax combined with the cognitive ability to organize related pieces of information about events and ideas in connected text.

(Berman 2008: 71)

Berman is particularly insistent that quantitative measures do not tap the full dimension of factors which contribute to proficient text production. She singles out the T-unit for particular criticism and does not make any reference to second language learning in her study but there are definite repercussions. Most importantly, it is not enough to register and quantify language use, it is necessary to explain it. This is obvious when dealing with native speakers, as by definition their command of the language is complete. However, the necessity to interpret data should also be applicable to advanced second language learners as they too can be expected to have crossed a threshold of knowledge. The hallmark of skilled performance should be how, not what, language is used.

The conclusion is not that quantitative measures have no value but that they have limitations. As ratios, they obscure the way that language is used in specific contexts to achieve specific goals. As units of measurement, they overrate the contribution of linguistic characteristics such as subordination. Research needs to dispense with preconceptions of complexity and focus on actual samples of language. To do this, it is necessary to examine individual writing texts from a learner corpus and evaluate how GC is affected. An argument for GC is built up by considering the context of production, the motivation for language use, understood as the linguistic choices available to the writer, and the skilful and accurate execution of the language. Where appropriate, native-speaker usage, as evidenced by primarily the British National Corpus (BNC), is taken into account in order to show the extent to which learner language adheres to or deviates from norms. While this paper now switches to a mainly qualitative approach, the issue of the contribution of quantitative measures to GC is taken up again in the final discussion section.

### 3. Complexity in context

The texts referred to are taken from the International Corpus of Learner English (ICLE) (Granger et al. 2002), a 2.5 million word collection of advanced essays written by undergraduates representing eleven first language groups. The ICLE is the most important learner corpus, based on a survey of the field (Pravec 2002), and the medium for extensive research activity (see, for example, Granger, 1998). The importance of the ICLE to this study is that it allows second language usage to be evaluated in a full context, i.e. a complete essay. This is often important in appreciating the grammar at a discourse level. To illustrate, consider the extract below from an essay on the theme of the conflict between technology and imagination. At this point the essay switches between complex and simple structures to reinforce the argument. First, a pseudo-cleft sentence (a *wh*-nominal clause with a copular complement) gives a rhetorical flourish by putting the rheme (basically, new information) at the end of the sentence.

What distinguishes human species from the rest of the animal world is its eternal longing for everything that is happier and greater. (ICLE-BG-SUN-0162.1)

Next, there are four noun phrases, each written on a new line, which are parallel to the predicate above.

Its longing for heaven.  
Its longing for freedom.  
Its longing for beauty.  
Its longing for love.

The repetition of elements that are not sentence constituents is a simple yet striking device which contrasts with the heavy pseudo-cleft. The last phrase *Its longing for love* acts as a climax, signalled through the /l/ alliteration and the switch to the monosyllabic noun, *love*. The next sequence is also grammatically incomplete, a non-embedded subordinate clause, but the tone is rather different.

Not that we have achieved that much.

Huddleston & Pullum (2002) note that the syntax of *not + that* clauses is problematic. The clause is non-embedded but *not* modifies a *that*-nominal. Strictly, the construction is subordinate in form but not in function. Non-embedded subordinate clauses, according to Biber et al. (1999), are a mark of spoken rather than written registers, so there is a drop in formality. Semantically, as Summers (2005) notes, *not + that* clauses serve to devalue the importance of the previous information, as can be witnessed in this excerpt from a cricket text in the BNC:

My prediction? Pakistan should make the semis, with a resurgent England and an improved India. Australia to win (or at least make the MCG final), with Pakistan the other likely finalist. Not that I'm putting money on any of them!

The last sentence shows a lack of confidence in the commentator's own predictions. The commentator first makes a forecast then effectively disowns responsibility for them. There is a hint of disingenuousness here, for why make bold assertions which are not well-founded? Perhaps the commentator wants to pre-warn the audience that this view may be fallible, hence saving face if the prediction does not go to plan. The *not + that* clause in the learner essay similarly downplays the argument for all this longing has not translated into enough results. A sense of bathos is felt, even humour. Interestingly, many of the BNC examples of *not + that* clauses, including the cricket text, also display humour so possibly this is part of the semantic prosody of the construction. The nominal relative which follows continues the

theme of under-achievement and the colloquial feel of *next-to-nothing* (citations in the BNC are mainly from speaking or dialogues in novels) maintains the less elevated tone:

We have even achieved what is next-to-nothing.

The repetition of the plural pronoun and *achieved* in the same perfect aspect reinforce the writer's position. The final two units of the essay consist only of a *to*-infinitive:

But in the meantime we continue to long. To dream. To achieve.

Again, the subordination is implicit rather than grammatically explicit. There is ambiguity over which verb licenses the infinitive. The most natural reading is probably that they both complement *continue*:

... [we continue to long] [and we continue to dream] [and we continue to achieve].

An alternative is a complex catenative structure with each verb in the sequence generating an object.

... [we continue [to long [to dream [to achieve]]]]

The meaning is that the achieving is part of a long, involved thought process. The chain of verbs is iconic for the successive prolongation of the predicate reinforces the idea of continuation. Syntactically, catenation could continue indefinitely so there is the sense of a never-ending process of hope and volition. A similar phenomenon is found in a lyric of Paul Simon (1966).

So, I'll continue to continue to pretend.

My life will never end.

Listening without the benefit of a song-sheet, *My life will never end* can be construed as an object of *pretend*, adding an interpretation of the self-delusion of immortality. The learner essay is a very different genre but it uses the same syntactical technique to bring the narrative to a conclusion which is effective not despite but because of the ambiguity. The ambiguity is indicative of the dual forces of progress and imagination as portrayed in the essay. They conflict with yet complement each other in a way that is both inspiring and disturbing. This argument is not based just on the last line of the essay, it is supported by evidence of skilled language use, in particular the adroit use of non-embedded subordinate clauses to provide subtle shifts in register and meaning.

Subordination features prominently in the discussion above and two important observations can be made. First, form and function do not always coincide for, as in the *Not that* sentence, subordination may not be syntactically dependent. This phenomenon is not registered in quantitative measures such as the T-unit. The presence of non-embedded structures would thus artificially depress statistics. Second, more crucially, subordination can only be appreciated when it is evaluated in context, ideally one which encompasses the entire essay.

Consideration of form alone, whether the nominal relative in a pseudo-cleft construction or a nominal *to*-infinitive, is not meaningful without providing a context-based rationale for the language use. Subordination acquires purpose in conditions which display it to the best linguistic advantage. There is no array of complex forms that automatically endow GC on a text. GC resides in how form and meaning interact and complement each other.

It is relatively simple to debunk the superiority of unit length. Below is the opening paragraph of an essay discussing the impact of unemployment on individuals and society:

You toil and beaver all through your life. Hardship and effort. Fatigue and exhaustion. And then it arrives out of the blue. You're thunderstruck, completely aghast. Nowadays, it just happens. Powerlessly, you stand in front of the factory gate watching the others passing by. It torments you. You can't believe it. You were not effective enough. Now you have had it. You were made redundant. (ICLE-GE-AUG-0040.3)

Orthographically, there are twelve sentences packed into this short paragraph. Many are minimal, e.g. *It torments you.* and two sequences, *Hardship and effort. Fatigue and exhaustion.* consist just of coordinated nouns. Only the one sentence *Powerlessly...*, is developed in the sense that it has subordination. Clearly, the syntax is minimalist, but this is by design for the structural simplicity of the text belies considerable sophistication of composition.

First note the predominance of the pronoun *you*, it occurs eight times, five times as first word in the sentence. The *you* is not spelled out but it must go beyond the reader of the essay to all members of society. The emphasis first brings home the warning of unemployment: *you* could be all and anybody, including the reader. The repetition also indicates an additional unpleasant side effect of unemployment, namely how self-absorbed it makes the ex-worker. Turning to the syntax, the staccato feel of the short sentences stresses the shock of unemployment, the *thunderstruck* effect. The accumulation of short sentences mirrors the way that everything seems to happen quickly and remorselessly in a crisis. The final sentence, *You were made redundant*, is the only one with a passive verb phrase. This is fitting for the unemployed are portrayed as passive victims of a process. Appropriately, this is a short passive because the agency behind unemployment is unknown and completely impersonal. Indeed, the remainder of the essay argues that perhaps the most frustrating thing about unemployment is its inhuman impartiality and indifference.

We are dealing with writing of the highest calibre here. Aside from the points made, the extract is error-free and it displays low-frequency and idiomatic lexis such as *beaver, out of the blue, have had it*. The force of the rhetoric is achieved not despite but because of the short sentences. The latter are a deliberate strategy to impose a very vivid impression on the reader. Aligning unit length with complexity misses the fact that the internal arrangement and combination of sentences are motivated by context. The linguistic realisation of the goals of production cannot be preconditioned or predicted; that is to say, common preconceptions of skilled writing often do not match to the actual achievements of writers. Thus, as shown, there is nothing complex *per se* about long/short units, everything depends on how, where and why they are used. GC cannot be reduced to a formula of constructions and text characteristics, it requires an understanding of the way that form and content interact in individual environments of use.

Texts differ on many dimensions, notably genre, but by definition the one unifying feature is context. It is therefore essential that complexity accounts for context. The major failing of quantitative measures is that they operate with tools that are abstract from the actual circumstances of production. Hence, they underestimate and under-appreciate the challenge of second language writing. Most damagingly, quantitative measures make presumptions about the relationship between language and GC which does not account for the freedom individual users have to choose and shape text. Writers select language through a mechanism of comparing different constructions and weighing up their effectiveness for the task in hand. A developed grammatical competence is signalled by the ability to appreciate distinctions such as register, frequency and style. Thus, GC unites richness of form and meaning in a way which resists classification schemes.

Nevertheless, quantitative data must have something to contribute to GC. By saying that quantitative measures are inadequate, there is the danger of throwing the baby out with the bathwater, so to speak. Whenever GC is related to language form, the presence, pervasiveness and representativeness of linguistic features will always demand comment. If such data become irrelevant, grammatical form is dangerously downgraded, and in the absence of form it is difficult to know what grammar is supposed to consist of. There can be no objection to quantitative measures *per se*, this would equate to rejecting form as a criterion in GC, an untenable position, but these measures could be better related to GC. Possibly, the methodologies used to date have been too limited, and the elaboration of GC awaits a wider battery of more powerful measures. Alternatively, the position here, quantitative measures

require interpretation and analysis at the level of individual sentences in specific contexts of use. If global trends and statistics are used to support localised language description in this way, quantitative and qualitative methodologies become complementary and non-exclusive. A triangulation of approaches is encouraging for, as the next and final section argues, it addresses the reliability problem that troubles an approach to GC which depends entirely on a personal reaction to the text.

#### 4. Discussion and conclusions

This paper has argued that GC results from how the writer selects and exploits language for the optimal means of expression. Since GC does not exist independently of writing, it cannot be pre-assigned to privileged notions of language form and structure. Language acquires GC from context and an essential part of the process of establishing GC is to consider the role of the language in developing a piece of writing. The move towards qualitative analysis does raise the question of reliability as it offers a more personalised response to writing. This involves some degree of judgment of factors such as effectiveness, style and appropriacy, leading to a charge of subjectivity. ‘No eye may see dispassionately’, as Peake has it in *Titus Groan* (1946/1992: 96). Quantitative measures are easier to defend as they are purely descriptive, not interpretative. As demonstrated, it is so easy to show the fallibility of measures such as the T-unit that this appeal to objectivity must be the only justification for purely quantitative methodologies. Reliability is crucial for sound linguistic analysis so a charge that alternatives to quantitative methods are unscientific must be countered.

First, it is important to appreciate that the qualitative approach taken relies on introspection rather than intuition. Itkonen (2008) makes the distinction between intuition and introspection that intuition is immediate and instinctive, one’s gut feeling, so to speak. Introspection on the other hand is deliberate and informed reflection on language use with the goal of understanding or manipulating it as a language event. Thus, introspection is based on intuition but it uses and develops it, for a purpose. As such, introspection can constitute a reliable methodology, whereas intuition cannot. The qualitative analysis in this study clearly goes beyond naked intuition for it examines the grammar in a way that incorporates considerations of the full context while appealing to corpus evidence to substantiate the claims made. Also, the analysis is completely transparent because the data, the ICLE, is available in its entirety for researchers and interested parties to examine. Furthermore, unlike with intuition, users could be trained in the analysis employed here, for example by exposure to learner corpora and manual parsing exercises which focus on identifying GC. It is not realistic to eliminate subjectivity from the approach completely but qualitative analysis can be sufficiently robust if it is applied in a consistent and informed way to the data.

Second, as stated earlier, quantitative results can balance and bolster a more subjective approach. Introspection, unlike intuition, is evidence-based and quantitative data supplies the raw material for investigations into and pronouncements on individual usage. To show how this might work in practice, a finding that a text includes a higher density of specific clause types than is the norm would be flagged in order for manual analysis to determine how significant this result was in context. Hence, qualitative analysis builds upon, rather than replaces, the quantitative findings, and they work in tandem to identify and appreciate GC. On a practical level, a dual focus also saves resources as the most interesting linguistic features can be identified and explored more efficiently. Indeed, with large corpora, it is difficult to see how research into GC would be viable without some degree of automatic analysis. Even the most steadfast advocates of qualitative approaches should admit that quantitative measures facilitate research and maximise its potential to be exploited. Certainly, a combination of both methods is the most attractive solution to reliability concerns.

As things stand, the emphasis on purely quantitative measures is too pronounced. A



number of disciplines use GC to mark and differentiate higher levels of performance but the uncritical acceptance of quantitative methods has deterred progress in defining valid stages and criteria of writing proficiency. Indeed, the increasing impact of technology makes quantitative measures ever more attractive. In this regard, it is a concern that major examination boards such as ETS and Cambridge ESOL are investigating the machine rating of writing tests, for while new technology has ready face validity and is often cost effective, such developments accentuate the privileged status accorded to quantitative data (see discussion in Rimmer, forthcoming). In this article it has been argued that human introspection adds a dimension to GC which cannot possibly be replicated or discarded. A case is made for a sounder notion of GC, i.e. an approach marked by a commitment to qualitative analysis.

We live in an age which demands accountability. This is most evident in high-stakes testing where candidates are often focused on obtaining target scores, perhaps regardless of how these scores translate into learning outcomes, and exam boards are under pressure to release statistics which justify the value of their tests. Naturally, language researchers also crave the illusion of quantifiable results in order to justify their investment of time and resources, as well as to provide a neat picture of accomplishment. In this environment, that which can be measured tends to accrue more credibility than that which cannot. Unfortunately, GC cannot be pinned down so easily and it is resistant to attempts to categorise it. The wrong reaction to this would be to dismiss GC as a variable on the grounds that it is too unstable and subjective. A more promising response, the line taken in this study, is to accept and appreciate GC on its own terms with a methodology that is sensitive to how GC is engineered in specific occasions of use.

## References

- Aronsson, M. (2003). On clefts and information structure in Swedish EFL writing. In Granger, S. & Petch-Tyson, S. (eds) *Extending the Scope of Corpus-Based Research: New Applications, New Challenges*. Amsterdam: Rodopi, 197-210.
- Berman, R. (2008). The psycholinguistics of developing text construction. *Journal of Child Language Acquisition* 35, 735-771.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Essex: Pearson Education.
- Brown, A. (2006). Candidate discourse in the revised IELTS speaking test. In McGovern, P. & Walsh, S. (eds) *IELTS Research Reports 6*. Canberra: IELTS, 71-90.
- Carter, R. & McCarthy, M. (1997). *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Carter, R. & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Granger, S. (ed.) (1998). *Learner English on Computer*. New York: Longman.
- Granger, S., Dagneaux, E. & Meunier, F. (2002). *International Corpus of Learner English*. Louvain: UCL Presses Universitaires de Louvain.
- Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, R. & Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Itkonen, E. (2008). Concerning the role of consciousness in linguistics. *Journal of Consciousness Studies* 15, 15-33.
- Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: how distinct? *Applied Linguistics* 29, 24-49.
- Kennedy, C. & Thorp, D. (2007). A corpus-based investigation of linguistic responses to an IELTS academic writing task. In Taylor, L. & Falvey, P. (eds) *Studies in Language Testing 19*. Cambridge: Cambridge University Press, 316-377.
- Lazarte, L. & Barry, S. (2008). Syntactic complexity and L2 academic immersion effects on readers' recall and pausing strategies for English and Spanish texts. *Language Learning* 58, 785-834.
- Matthews, P.H. (1981). *Syntax*. Cambridge: Cambridge University Press.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics* 24, 492-518.
- Peake, M. (1992). *Titus Groan*. London: Vantage. First published 1946.
- Pienemann, M. (1998). *Language Processing and Second Language Development*. Amsterdam: John

- Benjamins.
- Pinker, S. (2004). Clarifying the logical problem of language acquisition. *Journal of Child Language* 31, 949-953.
- Pravec, N. (2002). Survey of learner corpora. *ICAME Journal* 26, 81-114.
- Purpura, J. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English language*. Essex: Longman.
- Richards, J. & Schmidt, R. (2002). *Longman Dictionary of Language Teaching & Applied Linguistics*. Essex: Longman.
- Rimmer, W. (2008). Putting grammatical complexity in context. *Literacy* 42, 29-35.
- Rimmer, W. (forthcoming). Operationalising linguistic creativity. In Taylor, L. (ed.), *Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment*. Cambridge: Cambridge University Press.
- Sampson, G. (2003). The structure of children's writing: moving from spoken to adult written norms. In Granger, S. & Petch-Tyson, S. (eds) *Extending the Scope of Corpus-Based Research: New Applications, New Challenges*. Amsterdam: Rodopi, 176-193.
- Simon, P. (1966). Flowers never bend with the rainfall. In *Parsley, Sage, Rosemary and Thyme* (Audio CD). New York: Columbia Studios.
- Summers, D. (ed.) (2005). *Longman Dictionary of Contemporary English*. 4th edition. Harlow: Pearson Education.
- Towell, R. (2002). Relative degrees of fluency: a comparative case study of advanced learners of French. *IRAL* 40, 117-150.
- Willis, D. (2003). *Rules, Patterns and Words*. Cambridge: Cambridge University Press.
- 

Wayne Rimmer has taught English in Moldova, Russia, Germany, Thailand and the UK. He is currently writing a pedagogic grammar for Cambridge University Press. Email: w.rimmer@reading.ac.uk.