Marine Makhatadze

Perfectly Perfect Adverbs: A Learner Corpus Study of the Amplifier Collocations by Georgian Learners of English

Abstract

This study examines how adverbs of degree tend to collocate with particular words produced by Georgian Learners of English, both from a quantitative and qualitative perspective. Various collocation measures are used to show the strength of the bond between selected amplifiers (maximizers and boosters). The research is based on the contrastive corpus analysis of native speakers' corpus (Michigan Corpus of Upper-Level Student Papers) and Georgian learners' corpora (compiled by me in accordance with international learner corpus research guidelines). The aims of our research are: a) To investigate whether natives and Georgian Learners of English demonstrate similar patterns of saliency and agreement in their judgments of adverb + adjective collocations; b) To reveal the difference between the syntactic patterns and lexical meaning of the adverb + adjective collocations produced in written language by Georgian learners of English and native speakers?

It is hoped that in the light of the corpus-based evidence on the nature of collocations presented in this study, metalexicographers and learners will be informed about the gaps in terms of collocations acquisition and performance. Amplifier collocations might be expected to have a more explicit and prominent place in learner lexicographic research and the language teaching curriculum.

Keywords: learner corpus studies, amplifiers, collocations.

I believe the road to hell is paved with adverbs. They're like dandelions. If you fail to root it out the next day your lawn is totally, completely, and profligately covered with dandelions.

- Stephen King, *On Writing*

1. Introduction

Language production of non-native users of English continues to inspire research into the puzzle of nativelike selection of collocations. It is well established that phraseology is one of the aspects that unmistakably distinguishes native speakers of a language from L2 learners. As the written data of foreign language learners is an infallible resource for determining language competence, we started to compile a learner corpus of Georgian learners of English. Generally, learner corpora are the important lexicographic

"deposit" through which the linguistic internalisation of foreign or second language learners is observed. In our study, we decided to analyse how adverbs of degree tend to collocate with particular words produced by Georgian Learners of English, both from a quantitative and qualitative perspective. The aims of our paper are:

1. To investigate whether natives and Georgian Learners of English demonstrate similar patterns of saliency and agreement in their judgments of adverb + adjective collocations;
2. To reveal the difference between the syntactic patterns and lexical meaning of the adverb + adjective collocations produced in written language by Georgian learners of English and native speakers?

The present investigation concerns the most frequent amplifier collocations in corpora of novice academic English representing advanced learners (with Georgian as their L1) and native speakers of English.

The paper is structured as follows: after a review of relevant previous research and a presentation of material and method, amplifier collocations in learner corpora will be identified and discussed. Then follow several case studies of selected formulaic sequences, how they are used as regards their distribution, meanings and functions, that are either overused or underused by the learners before some concluding remarks are offered.

2. Some of the previous studies on amplifier collocations in learners of English

There is an abundance of phraseological sequences in the literature: idioms, lexical bundles, phrasal verbs, collocations, collostructures, etc. Collocations are the object of this study, and as such deserve an introduction and explanation. Collocations in general, and maximizer and booster collocations in particular, are important to study because they are often learned unconsciously by native English speakers. By studying them further, a better understanding can be gained of how non-native English speakers learn and use them (Kennedy 2003:481). There are many diverse definitions and approaches to collocations, but two main views have been particularly relevant in the literature (Nesselhauf, 2005). However, these are ready-made linguistic building blocks larger than words that are used (by the native speaker) as units of form and meaning. Collocations are combinations of words that occur together more often than would be expected by chance. There are three main approaches to understanding collocations:

the phraseological approach, the frequency-based approach and the significance-oriented approach, which is usually adopted by researchers in the field of pedagogy and/or lexicography (Herbst, 1996).

According to Cowie (1998), the phraseological approach sees collocations as fixed and semi-fixed expressions, meaning that the words in a collocation are often used together in a particular order and with little variation. This approach emphasises the importance of the relationship between the words in collocation and suggests that these relationships are often idiomatic or metaphorical. In other words, the meaning of the collocation is not simply a combination of the meanings of its individual words. Rather, the collocation as a whole has a specific meaning that is greater than the sum of its parts.

The frequency-based approach, on the other hand, views collocations as statistically significant patterns of word co-occurrence. The Neo-Firthian school gave rise to the statistically oriented approach or the frequency-based approach by Nesselhauf (2005). This approach is usually employed by researchers in the field of corpus linguistics and it highlights the importance of the frequency of occurrence of a particular word combination. Collocations that occur more frequently are considered stronger and more established than those that occur less frequently. Whichever approach is taken, corpus linguistics allows researchers to automatically extract large amounts of data, which can then can be analysed in terms of frequency or other collocational measures.

The present analysis is concerned specifically with the use of amplifiers. There are two major subcategories of amplifiers, namely, maximizers and boosters. Maximizers intensify the sense of an adjective. One can be, for example, *completely reckless*, *utterly exhausted*, *totally blank*, etc. Boosters, on the other hand, signify less than maximal intensity, for example, *very tired*, *particularly vulnerable*, *extremely unhealthy*, *heavily dependent*, *highly skilled*, etc.

Among corpus-based studies about amplifiers, based on a comparison of the characteristic modifier-adjective associations of native and non-native speakers in corpora, Lorenz (1999) suggested that German learners of English have a tendency to overuse particular modifiers and hyperbole. Another analysis of amplifier-adjective use is included in Biber, Johansson, Leech, Conrad, and Finegan's (1999) work, based on a large corpus of American and British spoken and written English. The description shows that *absolutely*, *very*, *bloody*, *damn*, *real*, *completely*, and *totally* are the most frequent amplifiers in British English conversation. In American English conversational genres the results and distribution is similar, with the exception of *bloody*, which is infrequent in American English. As for *extremely*, *highly*, *entirely*, *fully*, *incredibly*, etc. they occur repeatedly in both regional varieties, especially in written, academic genres.

Kennedy (2003: 472) retrieved 24 amplifiers and the words they modify from the British National Corpus, in order to explore collocational relationships between them. The amplifiers were selected largely because they are among the most frequent in the corpus and due to their frequent occurrence, in our research we will investigate these amplifiers (maximisers and boosters). Likewise, as mentioned above, in order to research the amplifier collocations in our learner corpus, frequency-based approach will be taken into consideration.

3. Material and Method

The following corpora form the core material for the present investigation: the Learner corpus of the Georgian learners of English (compiled by us) and the native speakers' corpus: Michigan Corpus of Upper-Level Student Papers. Each corpus contains student writing within a variety of academic disciplines. For the present purposes, only the domains of linguistics and literature have been investigated, and only texts written by students whose mother tongue is Georgian. Table 1 shows the size and composition of the corpora used.

Disciplines (Linguistics and Literature)	Number of words (tokens)
Learner corpus of the Georgian learners of English	62,555
MICUSP (Michigan Corpus) *In this case, only the English discipline Is chosen*	265,396

Table 1 Main corpora for the study

Our learner corpus has been annotated and extracted by means of AntConc (Antony, 2022) in order for searches to ignore material not produced by the student, such as linguistic examples, quotations and bibliographies.

Collocations associated with the 24 amplifiers listed in Table 2 are the source for our learner data analysis. The amplifiers, suggested by Kennedy (2003: 472) were selected largely because they are among the most frequent in the corpus.

Maximizer	Frequency	Booster	Frequency
fully	89	very	1,228
completely	86	really	476
entirely	69	particularly	219
absolutely	58	clearly	153
totally	58	highly	91
perfectly	44	very much	80
utterly	13	extremely	68
dead	8	badly	43
		heavily	41
		deeply	37
		greatly	33
		considerably	30
		severely	18
		terribly	13
		enormously	8
		incredibly	8

Table 2 The most frequent Amplifiers in the British National Corpus (per Million Words)

Note. From "Amplifier Collocations in the British National Corpus: Implications for English Language Teaching", by Kennedy, G. (2003). Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly*, *37*(3), 467–487. https://doi.org/10.2307/3588400

The statistical measure chosen to show the strength of the associations between amplifiers plus adjectives in the study was the mutual information (MI) measure. The higher the ratio, the stronger the association between the words. An MI score greater than 2 can be considered high enough to show a substantial association between two words.

3.1. Research participants

Corpus of Georgian learners of English has been set up with the aim of creating a tool for amplifier collocation research. The corpus consists of the academic essays written by Georgian learners (BA and MA students) of English at Ivane Javakhishvili Tbilisi State University, with the concentration of English Philology. The average age of the participants is 20 (median = 20) and 86% are females. The predominance of females is to be expected as with all humanities courses, thus, although the data may not appear well distributed, it is representative of the population found in the English Philology degree course. The participants are all Georgian native-speakers with Georgian native-speaker parents and they attended primary and secondary schools with Georgian as the medium of instruction. The average

number of years studying English at school is 8 (median = 8, IQR = 1), while the average number of years studying English at university is 3 (median = 3).

The majority are in their first term of full-time studies. An average essay length is 500 words. A typical first-term essay is somewhat shorter. The essays cover set topics of different types (reports, argumentative essays, narrative essays, etc.). They were written out of class, with a deadline of 1-2 weeks. Consent to participate and permission to use essays in the corpus were given in writing and students also completed a questionnaire providing information for the meta-information. Usually a learner corpus may contain learner-related variables (age, gender, and other information regarding the learner) and other metadata, such as L1, parents' L1, L3, and school year. These variables are collected because they may provide insights into any potential influences the learner may have been subject to (Granger, 1998a: 8).

3.2. Data collection procedures

Students were informed about the research, its aims and purpose. They were encouraged to enrol, although on an entirely voluntary basis. All essays were written without supervision or time constraints (apart from date deadlines), and with access to dictionaries, and written and electronic sources for facts. Taking into account the principles of representativeness and balance, we present the genre variety of the material. The material consists of texts of several styles, for example, journalistic, artistic, and official-business styles. Below are some of the essay topics that the study participants completed.

• How important is family for you? Is the family relationship the most enduring of all?

• Importance of make-believe games for children's development.

• Recall one memorable day from your childhood.

• "Marriages, like chemical unions, release upon dissolution packets of the energy locked up in their bonding" - John Updike "An orphaned swimming pool".

• Happiness is there in front of our eyes, but we don't see it. Miracles do happen.

• A country you have visited which has impressed you. Describe your visit.

• A movie/play you have seen which particularly impressed you.

Later, students were reminded to send the papers via e-mails, or hand in as a handwritten format. We removed the students' names and other means of identification, converted the texts to plain text format, standardised certain items and then annotated them automatically by means of TagAnt

(Antony, 2022). Also, the students filled in a questionnaire, answering questions about themselves, concerning their first language, parents' first language, previous studies, exposure to English, etc.

4. Corpus Analysis

The amplifier collocations, combinations constituted by adverbs + adjectives were extracted from our corpus via AntConc. AntConc (Anthony, 2022) is a free concordancer software developed by Laurence Anthony and it is widely used for the retrieval of concordances, n-grams, collocations, and keyword lists. The corpus was previously POS-tagged with CLAWS tagger (Garside & Smith, 1997) and the extraction of the combinations was based on the CLAWS5 tagset. The first step consisted in the identification of all the adverbs (tag = RB) and adjectives (tag = JJ) in the corpus by means of a simple tag search in AntConc. Then, manual sorting of the adjective concordances was carried out to highlight all the extracted adjectives preceded by adverbs (see figure 3).
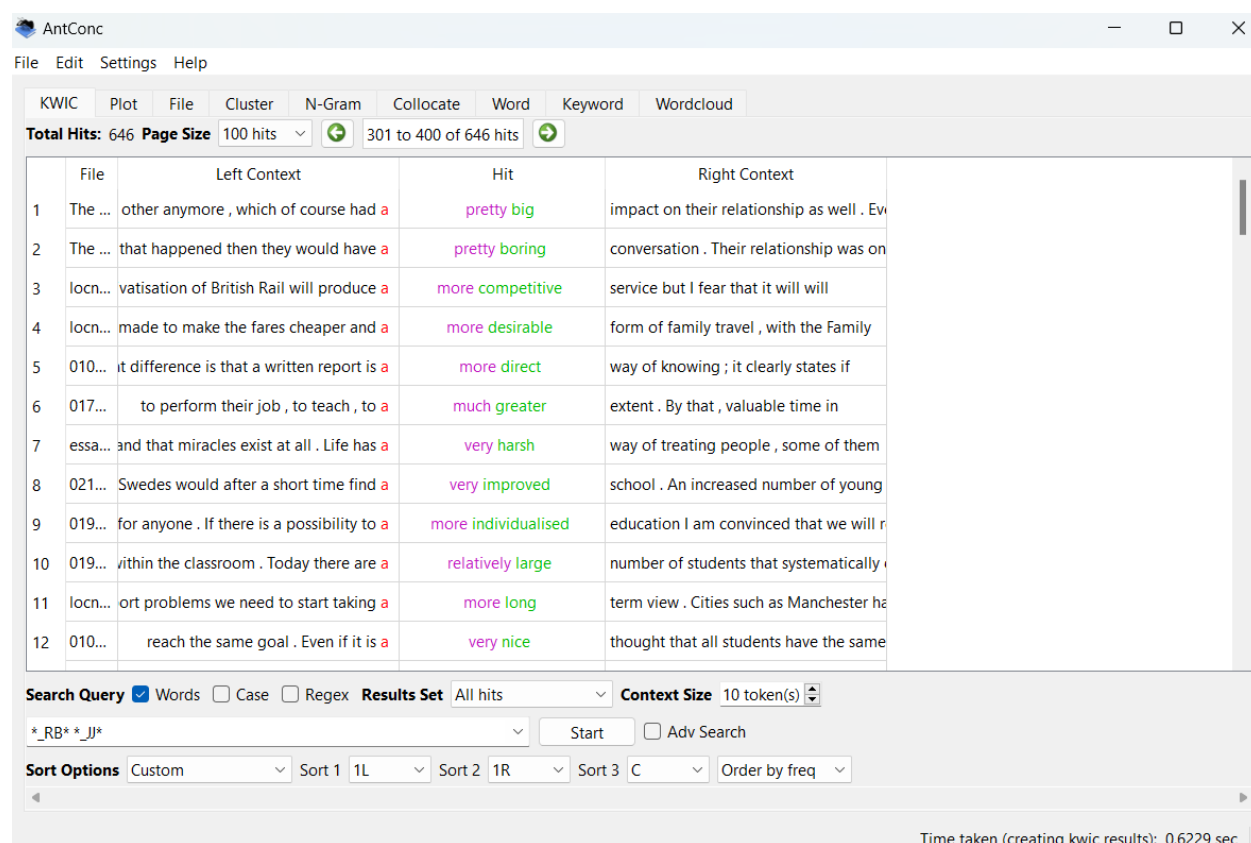

Figure 3: AntConc's sorting option set to - 1 (1L – left and 1R – right) highlighting adjectives (JJ) preceded by adverbs (RB).

Before we look specifically at amplifier collocations, we should first retrieve all the data representing adverb + adjective collocations. As the keyword in context shows, there are more than 600

examples of the given lexical pairs and here are some examples of the contexts: *I 'd want to emphasize that this is a **very special** and mysterious story; In my opinion , the author 's manner is **particularly intriguing**; This is because public transport is **less polluting** than personal car useage.*

N-gram analysis shows us adverb + adjective collocations ranked according to their frequency, range and normalized frequency (see the table 4).

	Type	Rank	Freq	Range
1	very important	1	23	19
2	most important	2	20	19
3	very good	3	12	10
4	how important	5	9	6
5	really important	9	7	5
6	very hard	12	6	5
7	very expensive	17	4	3
8	very nice	27	3	3
9	very valuable	27	3	3
10	Very few	27	3	3

Table 4. n-gram of adverb + adjective collocations in Georgian learners' corpus

Interestingly, the study shows that the underuse or overuse of collocations can be explained by lexical simplification strategies such as synonymy, paraphrasing or avoidance. This means that when learners are confronted with the use of a collocation (either in natural language or via an elicited task), they may rely on some lexical simplification strategies, such as that of synonymy, whereby they may produce a collocation whose collocate or node is a synonym of the correct lexical item.

Taking into consideration Kennedy's (2003: 472) list of the most frequent amplifiers in BNC corpus, we chose the following maximizers in this study: fully, completely, entirely, absolutely, totally, perfectly, utterly, dead. As for the boosters: very, really, particularly, clearly, highly, very much, extremely, enormously, considerably, deeply.

The results of the above-given amplifiers and their collocates in the corpora of Georgian Learners of English and native speakers are as follows:

Maximizer	Georgian Learners of English Corpus **Collocates**	Occurrences **Per 10K tokens**	MICUSP Michigan Corpus **Collocates**	Occurrences **Per 10K tokens**
completely	Completely different Completely private	3.90	Completely different, completely incorrect, completely undesirable completely uninteresting	3.73
entirely	x	x	entirely solitary, entirely successful, entirely fantastic, entirely unnoticeable entirely asexual	3.73
absolutely	absolutely crucial absolutely irrelevant absolutely new absolutely right	0.74	absolutely necessary absolutely homogenized	0.30
totally	Totally dependent Totally different	0.72	totally numb totally absent, totally fucked up	2.07
perfectly	perfectly unimportant perfectly invisible	0.32	perfectly content perfectly acceptable	1.58
utterly	x	x	utterly insane, utterly terrified, utterly unable	2.07

Table 5. Maximizer adverbs and their collocates ranked according to occurrences per 10K tokens

As for the boosters, the results given in the table highlight the contrast between native and non-native speakers of English. (See Table 6).

Booster	Georgian Learners of English Corpus **Collocates**	Occurrences **Per 10K tokens**	MICUSP Michigan Corpus **Collocates**	Occurrences **Per 10K tokens**

Very	Very important, Very hard, Very good, Very much, Very difficult	22.56	very powerful very human-like very different …	9.45
Really	really important, really precious really bad, really beneficial really essential, really happy really hard, really interesting really sensitive, really unpleasant really warm	3.25	really lovely really funny really tight	0.18
Particularl	particularly harmful particularly intriguing	0.36	particularly vibra particularly felicit particularly significa particularly interested	1.77
clearly	x	X	clearly definable clearly powerful clearly unnerved	2.86
highly	Highly prioritized Highly respectful	0.36	highly activated highly expressed highly conserved highly developed highly favored	13.41
extremely	Extremely inefficient	0.18	extremely illuminating extremely feminine extremely self-serving extremely magnanimous extremely brief extremely concerned	3.73

Table 6. Occurrences per 10K tokens for boosters according to corpora of American and Georgian students.

We observe that Georgian learners use significantly fewer –ly adverbs than native speakers in terms of types and tokens. This may result in a lack of diversification of linguistic items which involves less nativelike proficiency and fluency. This also contributes to the "*foreignsoundingness*" of learners, even those at intermediate and advanced levels of proficiency. When comparing the collocates of the maximizers and boosters, the results show that in most cases, the Georgian students do not use more collocations for each maximizer than the Americans do (however, exceptions are *very* and *really*).

5. Discussion

Besides the strength of the bonding between amplifier collocates, more general semantic and grammatical characteristics can be involved in the collocations, as these examples reveal: **completely** tends to be associated with adjectives that are used in a neutral connotation for Georgian students, whereas, American students use it with rather negative semantic associations, i.e. the adjectives have with negative prefixes ***un-***, ***in-***; Another example reveals that usage of **totally** tends to have mainly negative associations for Georgian learners and American students; Surprisingly, the amplifier **perfectly** has exclusively positive associations in native speakers' papers, whereas Georgian learners' usage of this word has somewhat negative associations in the contexts. **Extremely** tends to be associated especially with adjectives that have positive associations in Michigan Corpus, on the other hand, Georgian students used the adjectives with negative associations.

The amplifier collocations described in this study reveal a minuscule part of the learning that is necessary in order to become a fluent user of English. A substantial part of linguistic competence appears to be based on a huge store of collocations and their associations.

One of our research questions and aims is to use the learner corpus data for lexicographical purposes. In modern metalexicographic theory significant place is occupied by text graphs (the data boxes), which are quite a common phenomenon in modern lexicography. However, relatively less attention is paid to their representation in the microstructural fields of the dictionary article, what is the purpose and purpose of these textual elements and what kind of information should they reflect? Dictionaries offer a variety of data in text graphs, for example, in some cases they act as a kind of guide for analyzing the meaning of a word, sometimes they provide information on the limitations of the range of use of a word, pronunciation and register. One of the most important paradigmatic changes in modern metalexicography lies in the understanding that different types of microstructures allow lexicographers to present lexical items in a way that best suits the needs and reference skills of their

target users. The user's perspective plays a dominant role in the planning and final compilation of dictionaries.

This also applies to data distribution and data presentation in any given dictionary. In this sense, information about amplifier collocations, their usage peculiarities, which is best represented in the learner corpus, should be analysed and used by lexicographers in order to satisfy learners' needs. For example, when defining the particular adjective, the lexicographer can indicate the usage note in the lexicographical data box, where the user will get information about how to use the specific maximizers and boosters as the collocates.

6. Conclusion

The present study has looked into the amplifier collocations. The results show that there are indeed some differences in the way Georgian Learners of English and Americans use modifying adverbs, although they are not vast. In terms of frequency, there is only one maximizer and a booster that stands out, namely *completely* and *really*. The normalized frequency number shows that they are used more frequently in the Georgian learners' texts compared to the Americans'. The amplifiers that were not used at all in our Georgian learner corpus were *clearly*, *entirely*, *utterly*. Besides the strength of the bonding between amplifier collocates, more general semantic and grammatical characteristics can be involved in the collocations as well. It is hoped that in the light of the corpus-based evidence on the nature of collocations presented in this study, will inform metalexicographers and learners about the gaps in terms of collocations acquisition and performance. Amplifier collocations might be expected to have a more explicit and prominent place in learner lexicographic research.

References

Anthony, L. (2022). *AntConc (Version 4.2.0) [Computer Software]*. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2022). *TagAnt (Version 2.0.5) [Computer Software]*. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. & Quirk, R. (1999). *Longman grammar of spoken and written English*. Longman.

Cowie, A. (1998). *Phraseology: Theory, analysis, and applications*. Oxford University Press.

Granger, S. (2012b). Learner corpora. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3235–3242). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0669

Granger, S. (2017). Learner corpora in foreign language education. In S. Thorne, & S. May (Eds.), *Language and technology. Encyclopedia of language and education* (3rd ed.), (pp. 1–14). Springer International. https://doi.org/10.1007/978-3-319- 02237-6_33

Granger, S. 1998. Prefabricated patterns in advanced EFL writing. In Cowie, A.P. (ed.), *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, 145–160.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. https://doi.org/10.1111/j.1473- 4192.1994.tb00065.x

Herbst, T. (1996). What are collocations: Sandy beaches or false teeth? *English Studies*, 77(4), 379–393. https://doi.org/10.1080/00138389608599038

Kennedy, G.D. (2003). Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly, 37*, 467-487.

Lorenz, G. (1999). *Adjective intensification - Learners versus native speakers: A corpus study of argumentative writing*. Rodopi.

Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan. https://micusp.elicorpora.info/main

Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam/New York: John Benjamins.

Author's email: *marine_makhatadze@yahoo.com*

Author's Biographical Data

Marine Makhatadze is a PhD student at Ivane Javakhishvili Tbilisi State University, Faculty of Humanities. She teaches English at the Department of English Philology (the courses include Metalexicography and Learner Lexicography). She is interested in lexicography and corpus linguistics and is currently researching corpus-based learners' and native speakers' use of recurrent word combinations.