A Corpus-based Study on the Use of Contractions by EFL Learners in Argumentative Essays

M. Pınar Babanoğlu

Faculty of Arts and Sciences, Kahramanmaraş Sütçü İmam University, Avşar Campus, 46100 Kahramanmaraş Turkey

E-mail: pinarbab@hotmail.com

Abstract

Contraction forms in English are mostly occur in speech and informal writing and they are generally avoided in formal writing types such as academic prose, business reports and journal articles, therefore, most teachers discourage their use in academic essays (Biber, Johansonn, Leech, Conrad and Finegan 1999). Contractions in English have two types; negative contractions (isn't, haven't, doesn't) and verb contractions (I'm, they've, that's). This corpus based study attempts to investigate contraction usage in learner and native English speaker essays. Major goal is to examine whether learners consider essay writing rules in respect of contractions which are accepted inappropriate for academic prose style. Five corpora, three learner and two native English, were utilized in order to analyze verb and not-contraction forms. Frequency calculations of contraction forms in each corpus compared via log-likelihood measurement for statistical significance. Results revealed that learners use considerably more contraction forms, especially negative ones, than native English students in their argumentative essays.

Keywords: English for Academic Purposes, learner corpus, contractions

1. Introduction

Academic writing refers to writing in a particular style with a certain set of rules and patterns for a particular purpose. Main aim is to express a central point related to an argument structure with a formal, standard written language to inform a certain audience group including a community of researchers, lecturers, students, etc. Writing is essential for all students in higher education and it is a process that starts from understanding the task, planning and writing drafts, then goes on to the final text (Gillet, Hammond and Martala, 2009). Myles (2002) points out that 'the ability to write well is not a naturally acquired skill, it is usually learned or culturally transmitted as a set of practices in formal instructional settings or other environments (p.A1)' hence, writing skills must be practiced and learned through experience. Academic essays are written in formal English and are more complex than more informal writing or conversation, i.e. they have more longer words with more grammatical complexity including subordinate clauses and passives as well as it uses more noun-based phrases than verb-based phrases. In addition, some structures are avoided in formal writing such as colloquial words and expressions rhetorical questions and contractions like 'you're' or 'doesn't/don't'. Contractions are mostly in speech and informal writing and they are generally avoided in academic prose, business reports and journal articles (Biber, et.al.,1999) and are used mostly in speech and informal writing such as fiction. Consequently, while contractions can be very useful in written English, experts caution against the use of contractions in formal communication since they tend to add a light and informal tone to writing and they are often inappropriate for academic research papers, technical writing, business presentations, and other types of official correspondence. Although contractions have a certain linguistics identity in genres, contracted words such as 'don't', 'can't', 'shouldn't' are informal and should not normally be used when writing in an academic context (unless they are quotations which cannot be changed) (Gillet, et.al., 2009, p.96). However, some types of text such as fictional stories or novels, dialogue, or personal letters or emails, can benefit from the inclusion of contractions in order to create a more informal and/or conversational tone. On the other hand, although the contractions are strictly avoided for more formal prose types, yet they have a crucial role for other fields. Some people have an idea that contractions represent an efficient way of communicating in the era of text-messages via communication devices or social media and they also save authors from using every single letter of every single word (Tepper, 2014).

 Gilquin, Granger and Paquot (2007) emphasize that 'The analysis of learner corpus data and their comparison with data from native corpora have highlighted a number of problems which non-native learners experience when writing academic essays, e.g. lack of register awareness, phraseological infelicities, semantic misuse' (p.1). Thus, learner corpora studies may be beneficial to undercover certain problems of learners as learner corpora contain learners' L2 data and serve a potential for EAP studies. Gilquin & Paquot (2007) state that many learners use features which are more typical of speech than of writing, which give their essays an overly oral tone and this may be problematic for learners to conduct stylistically appropriate tone in their academic writing. For instance, Babanoğlu (2014) studied pragmatic markers in learner corpus and found that learners use oral features in their essays. According to Aijmer and

Strensöm (2004) learners may overuse or underuse certain structures in comparison with native speakers and therefore sound non-native in their L2 written products. Kilimci (2009) examined linking adverbials in learner corpus and found excessive overuse by all learners. With this motivation, this corpus based study investigates the contraction use of learners and native English speakers in their argumentative essays within learner and native English speaker corpora. Research questions are posed as follows:

1. Is there a statistical difference between native English speakers and learners in the use of contractions in their argumentative essays?

2. Is there a statistical difference between different learner groups in the use of contractions in their argumentative essays?

1.1 Theoretical Background

1.1.1 Contractions

Contractions in English are two classes; verb contractions (e.g. I'm, they're) and not-contractions (e.g. haven't, isn't) Verb contractions are composed of primary verbs (operators) 'be' and 'have' as well as modal verbs such as 'will' and 'would' (Biber, et.al. 1999, p.1129):

be	am- 'm	are- 're	is- 's
have	have- 've	has-'s	had-'d
Modals	will- 'll	would-'d	

Verb contraction can also be attached to –wh words, that and there like; that's, where's, there're. Not-contractions occur where negative particle 'not' is reduced to 'n't' losing its stress and attached to a preceding operator (main/auxiliary verb and modal verb):

are not	*aren't*	is not	*isn't*		
do not	*don't*	does not	*doesn't*	did not	*didn't*
have not	*haven't*	has not	*hasn't*	had not	*hadn't*
was not	*wasn't*	were not	*weren't*		
cannot	*can't*	could not	*couldn't*		
should not	*shouldn't*	would not	*wouldn't*	will not	*won't*

1.1.2 Studies on Contractions

In the literature, there is a gap in terms of research on contractions as a linguistic item as well as its usage and limited number of studies have been carried out so far. A large scale corpus research on contraction use in different registers Biber et.al. (1999) conducted an in-depth corpus study with a wide corpus, Longman Grammar of Spoken and Written English (LGSWE) which includes over 40 million words, and comprise of four kinds of text such as conversation, fiction, news and academic. In the study, verb and not-contractions were also investigated in four types texts corpora and Biber et.al. (1999) found that both verb and not-contractions had the highest frequency in conversation and fiction texts whereas they are at low rates in news and academic texts. That is, order of frequency levels of contractions is conversation, fiction, news and academic texts, which means contractions are mostly frequent in spoken registers and informal written registers and rare in formal written registers. Perez (2013) and Gonzalez (2007) studied negative contractions with a corpus data and confirmed Biber et.al.'s (1999) some of results about negative contractions in written and spoken register. In another corpus study, Olohan (2003) examined contractions in Translational English Corpus and found significant differences between English literary translation and contemporary literary English writing both in variety and frequency.

With regard to learner writing, Shaw & Liu (1998) investigated second language writing in a two-stage study with a group of learners from more than ten different native language (Arabic, Chinese, Persian, Turkish, Japanese, ..) in order to analyze the development of learners' writing in target language. They found that learners tend to write in too spoken tone although they speak in a register too written. In their study, contractions were one of their linguistic items to investigate in learners' writing development and learners who mostly use contractions were respectively Japanese, Arabic and German learners whereas Turkish learners used contractions at a very low frequency in essays (Shaw & Liu, 1998, p. 250).

2. Methodology

In the present study, Contrastive Interlanguage Analysis (CIA) (Granger, 2002) is regarded as the corpus methodology which is based on the statistical comparison of :

L1 vs. L2: Native vs. Non-native groups frequency comparison

L2 vs. L2: Frequency comparison among different Non-native groups

Specifically, frequencies of contractions are compared between learner corpora and native English corpora as well as learner corpora each other. Native and non-native speaker (learner) comparisons can highlight a range of features of non-nativeness in learner writing and speech, i.e. not only errors, but also instances of under and over representation of words, phrases and structures (Granger, 2002).

2.1 Data

In the study, five corpora, two native English and three non-native English used to investigate contraction types:

Two Native English Corpora

BAWE: British Academic Written English; a formal academic corpus Phd, Ma Theses, articles, projects, reports

LOCNESS: Louavin Corpus of Native English Students; essays of university students

Three NON-Native English Corpora (Learner Corpora)

Three Sub corpus from ICLE (International Corpus of Learner English), which is a large learner corpus of 3.7 million words, comprised of argumentative essays of learners of English from 16 language backgrounds, were utilized:

Turkish sub-corpus (TICLE):   argumentative essays of Turkish learners English

German sub-corpus(GICLE):   argumentative essays of German learners of English

Japanese sub-corpus (JPICLE):   argumentative essays of Japanese of learners of English

Table 1. shows the size of each corpora used in the study in respect of number of total words and texts:

Table 1. Distribution of five corpora in the study

Corpora	Number of texts	Number of words
LOCNESS (L1 English)	175	162.358
BAWE (L1 English)	80	256.151
TICLE (L2 English)	280	223.960
GICLE (L2 English)	265	256.151
JPICLE (L2 English)	264	227.147

2.2 Instruments

In order to analyze contractions within five corpora, online Sketchengine software (Kilgarriff, et.al., 2004) was used in the study. All verb and not contraction types used learners in their argumentative essays in native English and non-native English corpora were identified and calculated for their frequency. Afterwards, log-likelihood (LL) measurement, which is a very useful statistical method to compare corpora calculating the number of words in each corpus and frequency of the linguistic item, was applied to find out statistical differences among groups (http://ucrel.lancs.ac.uk/llwizard.html).

2.3 Data Analysis

The analysis method of the study is based on CIA methodology as stated above, that is applied due to comparisons of corpora were done as follows:

1. L1 vs. L2: Native vs. Non native comparison of contraction frequencies by log-likelihood ratio:

BAWE vs. TICLE   LOCNESS vs. TICLE

BAWE vs. GICLE   LOCNESS vs. TICLE

BAWE vs. JPICLE   LOCNESS vs. JPICLE

2. L2 vs. L2: Contraction frequencies by Log-likelihood ratio among Non-native groups:

TICLE vs. GPICLE

TICLE vs. JPICLE

GICLE vs. JPICLE

Log-likelihood measurement enables to reveal the possibility of contractions' frequency differences between two corpora, indicating an overuse or underuse in one corpus as a sign of statistical significance.

3. Results and Discussion

Initial analysis has been made via frequency calculation of all contraction types in each native and non-native corpus. Figure 1 shows overall frequencies in five corpora indicating a clear frequency difference in contraction use between native English and non-native English groups. In general, overall frequency of contractions is higher in non-native group in which contractions mostly used by Turkish learners of English.
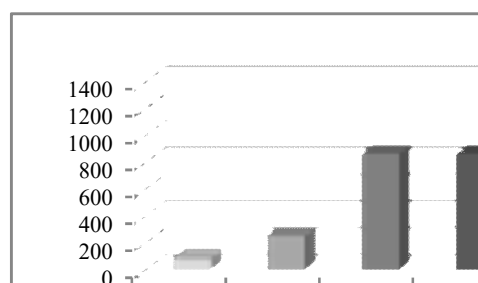
Figure 1. Overall frequency of contractions in five corpora

When verb and not contractions are calculated separately, the main quantitative density is in not contractions in both native and non-native English groups. Verb contractions have at the lowest frequency levels whereas not contractions are at higher levels when compared to verb contractions in both groups. Although not contractions are more frequent than verb contractions in native English corpora, they are far more in non-native English groups. The percentages of contraction use in all groups are given with type-token ratio in Table 2:

Table 2. Type/token ratio of contractions in five corpora

Corpora	Type/Token ratio	
	Verb Contractions	Not Contractions
BAWE	0.89	4.2
LOCNESS	1.74	20.7
TICLE	0.67	40.1
GICLE	1.89	39.0
JPICLE	0.55	49.3

In Table 2, Type-token ratio of contractions shows that not-contractions have a higher level of usage in all learner corpora and LOCNESS native English corpus. Japanese learners used almost half (49.3%) of negative forms as contraction. Similarly, German and Turkish learners also used contractions as negative forms at high rates as 39% and 40.1% whereas native English speakers used them less than learners, only students in LOCNESS corpus used negative forms as contractions at 20.7% rate which is half of learners. BAWE corpus, which contains more academic texts like theses, reports, etc., has the lowest frequency level of contractions.

In order to understand whether these frequency differences are statistically significant or not, total frequencies of contractions were compared by log-likelihood (LL) ratio (Table 3.).

Table 3. Log-likelihood ratio of overall frequency of contractions among learner and native corpora

L1 vs. L2	Verb Contractions	Not Contractions
	LL value	LL Value
Learner vs. Native English	+5.62	+2301.15

*$p < 0.05$ (critical value: 3.84) + indicates overuse in the first corpus relative to the second corpus, - indicates underuse in the first corpus relative to the second corpus

Log-likelihood results show that there is significant difference between learner and native English speakers in the use of contractions. Verb contractions were used slightly more by learners than native speakers of English with +5.62 log-likelihood values which means there is an overuse in the first corpus as it is higher than critical value (3.84). On the other hand, not contractions were significantly overused by learners when compared to native speakers with a high rate of +2301.15 log-likelihood value. Therefore, below the tables not-contractions are presented in detail.

Not-contraction forms *don't/doesn't/didn't* forms are presented in Figure 2. in which they were mostly used by Japanese learners, then respectively Turkish and German learners.
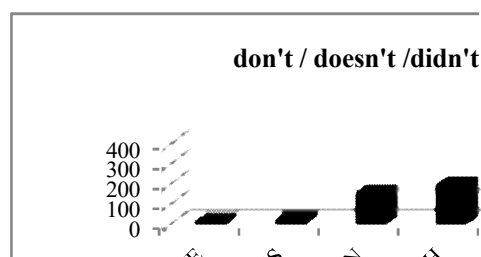
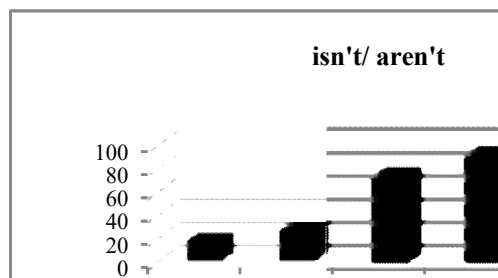Figure 2. Frequency of don't/doesn't/didn't in five corpora

Figure 3. Frequency of isn't/aren't in five corpora

Since the frequencies of not-contraction types indicated difference in frequency among learners and native English speakers, log-likelihood ratio of each group frequencies has been applied to clarify such difference is statistically significant or not. Frequencies of each not-contraction item were compared among groups to find out log-likelihood values in Table 4. and statistical significance is revealed in terms of frequency differences in negative contraction items.

Table 4. Log-likelihood values of all not-contraction forms in learner vs. native English corpora

L1 vs. L2	don't/doesn't/ didn't	isn't/ aren't	can't/couldn't	haven't/hasn't hadn't	wasn't weren't	shouldn't/ wouldn't/won't
TICLE vs. LOCNESS	+83.6*	+17.9*	+50.2*	+10.6*	-4.0*	+27.0*
TICLE vs. BAWE	+395.9*	+61.4*	+207.1*	+33.3*	+1.0	+134.3*
GICLE vs. LOCNESS	+52.5*	+22.4*	+22.9*	+18.2*	-5.4*	+33.5*
GICLE vs. BAWE	+328.3*	+37.7*	+143.8*	+47.6*	+33.7*	+150.6*
JPICLE vs. LOCNESS	+301.3*	+3.6	+160.5*	+14.1*	-3.3	+1.8
JPICLE vs. BAWE	+835.4*	27.5*	+419.5*	+39.8*	+1.4	+57.6*

*$p < 0.05$ (critical value: 3.84), + indicates overuse in the first corpus relative to the second corpus,-indicates underuse in the first corpus relative to the second corpus

Almost all not-contraction forms were significantly overused by learners against native speakers due to the fact that the higher the log-likelihood value the more significant the different is. German learners overused all forms against two native English corpora as well as Turkish learners (except for *wasn't/weren't*). Japanese learners overused three some forms but they used some of them equally as native speakers. As CIA (Granger, 2002) suggests for corpus comparisons methodology, frequency of negative contraction forms in learner corpora were compared each other via log-likelihood (LL) measurement (L2 vs. L2).

Table 5. Log-likelihood ratio of overall frequency of contractions among learner corpora

L2 vs.L1	Verb contractions LL value	Not contractions LL value
TICLE vs. GICLE	-25.2*	+1.81
TICLE vs. JPICLE	+0.2	-72.0*
GICLE vs. JPICLE	+30.4*	-100.0*

*p < 0.05 (critical value: 3.84), + indicates overuse in the first corpus relative
to the second corpus, -indicates underuse in the first corpus relative to the second corpus

In Table 5, when frequencies of all contraction forms were compared via log-likelihood measurement, it can be seen that there are some certain statistically significant differences between learner groups as well. For example, Turkish learners used less verb contraction forms than German learners and less not-contraction forms than Japanese learners. German learners overused verb contractions but underused not-contractions than Japanese learners. Namely, Japanese learners used less verb contractions than Germans but after all they overused not-contractions than all other learner groups whereas Germans overused verb contraction than all other groups.

4. Conclusion

Creating a formal writing voice in essay writing, certain rules are indispensible such as not to use first personal pronouns (I, me or we), slang or everyday speech words (yeah, cool, okay, kind of,..). Within this context, contraction forms in English are generally avoided in formal writing in order to establish a more formal tone in, for example, essays, study reports, theses. It is clear that, avoiding may be problematic not only for native English speaker university students but also for non-native English (learner) university students. According to the study results,

According to the results, in respect of first research question of the present research, contractions (mainly negative/not-contractions) are significantly overused in non-native English students' essays when compared to native English students' essays and academic texts. In Turkish language, for example, negation is set by two basic types as negation affixes (-**ma**, -**sız**) attached to verb stems and negation with words such as **hayır** (no) or **yok** (no/not) (İlhan, 2005). For instance, 'ödevi yap**ma**dım' (I **did not do** the homework) –**ma** is simple past first person negation and equals to **did not** and in 'bugün ödev **yok**' 'there is **no** homework today', and **yok** is a negation word which means **no** in English, namely, contraction of these forms is not possible. In German, 'nicht' and 'kein' words are used for negation and contraction is not available. In Japanese, standard negation is made by -**na** suffix directly attached to verb stem (Nyberg, 2012) which is not contracted as well. These language specific factors may be reason for Turkish, German and Japanese learners overuse contractions in their essays against native English speakers. Considering the fact that some forms which are peculiar to conversation or informal writing such as contractions give an informal tone to formal writing or academic prose types, the present study confirms the suggestions of some of past research (Ajmer and Strensöm, 2004; Gilquin, Granger and Paquot, 2007; Gilquin and Paquot, 2007; Kilimci, 2009; Babanoğlu, 2014) which state that learners may have problems in establishing an appropriate tone in formal writing.

In order to seek an answer for the second research question of the study, when learner groups compared each other, German learners mostly used verb contractions and Japanese learners overused not-contraction whereas Turkish learners are the only group that used all contraction forms less than other learner groups as Shaw & Liu (1998). The contrast between the performances in the use of contractions among learners may due to several factors such as L2 instruction on writing in target language or cross-linguistic differences.

Outcomes of this study can be regarded as an indicator for EAP methodology to help learners to conduct stylistically appropriate tone in their L2 writing. Consciousness rising on the spoken and written register differences, also on cross-linguistic differences, highlighting the relevant L2 writing instructions may support learners in order to achieve the proper tone for their academic writing. For further research, the number of learner corpora can be enhanced to see the use of contractions by different L2 groups from different native languages in order to generalize the fact that learners have trouble with using contractions in essays.

References

Aijmer, K. & Strensöm, A.B. (Eds.), (2004). Discourse Patterns in Spoken and Written Corpora. Philadelphia: John Benjamins.

Babanoğlu, M.P. (2014). A corpus-based study on the use of pragmatic markers as speechlike features in Turkish EFL learners' argumentative essays. *Procedia - Social and Behavioral Sciences*, 136, 186-193. Doi:10.1016/j.sbspro.2014.05.312

Biber, D., Johansonn, S., Leech,G. Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Gillet, A., Hammond, A. & Martala, M. (2009). *Successful Academic Writing.* University of Herefordshire, Pearson Education Limited.

Gilquin, G., Granger, S. & Paquot, M. (2007). Learner corpora: the missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-355. http://dx.doi.org/10.1016/j.jeap.2007.09.007.

Gilquin, G.& Paquot, M. (2007). Spoken Features in Learner Academic Writing: Identification, Explanation and Solution. Proceedings of the 4th Corpus Linguistics Conference, University of Birmingham, 27–30 July 2007.

Gonzales, M.P. (2007). *Uncontracted negatives and negative contractions*. Unpublished Doctoral Dissertation, Universtade de Santiago de Compostela, Spain.

Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, S. Petch-Tyson & J. Hulstijn (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam & Philadelphia: John Benjamins.

İlhan, N. (2005). Türkçede olumsuzluk. *Karaman Dil Kültür ve Sanat Dergisi*, 271-279.

Kilgarriff, A., Rychly, P., Smrz, P. & Tugwell, D. (2004). *The Sketch Engine*. Proc EURALEX 2004, Lorient, France; 105–116.

Kilimci A. (2009). Negotiation of meaning in L2 academic writing. In Socarrás, Gilda M. , Eds., *Philological explorations*, pp.47-58, Atina: ATINER Athens Institute for Education and Research Studies Press Ltd. [Online] Available at:
https://www.academia.edu/2970580/Kilimci_A._2009_._Negotiation_Of_Meaning_In_L2_Academic_Writing._In_G._
Socarras_Ed_Proceedings_of_1st_International_Conference_on_Literature_Languages_and_Linguistics._Athens_ATI
NER

Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. TESL-EJ, 6(2). [Online] Available: http://tesl-ej.org/ej22/a1.html.

Nyberg, J. (2012). *Negation in Japanese*. Unpublished Doctoral Dissertation. Stockholm University, Sweden.

Olohan, M. (2003). How frequent are the contractions? A study of contractions in the Translational English Corpus. *Target*, 15(1), 59-89.

Perez, V. (2013). Operator and negative contraction in spoken British English: a change in progress. *In* B. Aarts, J. Close, G. Leech and S. Wallis, (Eds) *The Verb phrase in English: investigating recent language change with corpora* pp:256-285. Cambridge University Press. [Online] Available at: https://www.cambridge.org/core/books/the-verb-phrase-in-english/7A9E822B172BA29A2A73660B96C13003.

Shaw, P.& Liu, E.T. (1998). What develops in the development of second-language writing? *Applied Linguistics*, 9(2), 225-254.

Tepper, N. (2014). *9 English contractions you should never use and why.* [Online] Available at: ttps://www.kibin.com/grammar-rules-blog/9-english-contractions-never-use (23.05.2016).