

IEP Listening Tests: An Exploratory Study

Deirdre J. Derrick

Northern Arizona University

Abstract

This exploratory study used a corpus methodology to explore select lexis and grammatical features in the placement and exist listening tests of a North American Intensive English Program (IEP). The objective of the study was to see if there were differences among the eight registers on the tests. After a corpus of the listening texts was created, programs were written using the computer language Perl to find certain features in the texts. Once these features were collected, a series of one-way ANOVAs were run to look for statistically significant differences among the registers. No significant differences were found, leaving the door open for future research. Implications of this study for the IEP are that differences in the tests may exist in something other than morpho-syntactic complexity, at least as measured by this programs written for this study.

Keywords: corpus methodology, Intensive English Programs, listening tests

IEP listening tests: An exploratory study

Background

Corpus methodology can be useful for comparing actual language with the language learners are taught. Despite this useful application, corpus research has mainly been used to guide dictionary development (Burton, 2012). This becomes apparent when ESL/EFL textbooks are examined. Researchers have found inconsistencies in how spoken language is presented in textbooks and how spoken language and interaction occur. For example, Carter (1998) examined textbooks qualitatively to compare them with the CANCODE corpus, focusing on three-part interactions, vague language, and ellipses. He found that language from the CANCODE corpus “broke” several of the grammar rules presented in the textbooks and argued for the importance of teaching learners “real” spoken language.

Related to the use of corpora to inform textbook materials development is the use of corpora to inform ESL/EFL listening test development. This use appears to be more common in large-scale, high-stakes test development. The TOEFL iBT incorporated corpus research into its design. The Michigan Corpus of Academic Spoken English (MICASE) has been used to develop listening tests, and the University of Cambridge Local Examinations Syndicate (UCLES) has been active in corpus development and uses corpus for its English for Specific Purposes (ESP) tests (McNamara, Hill, & May, 2002).

However, beyond its use for the development of high-stakes English tests, corpus methodology does not appear to be commonly or frequently used for lower-stakes tests, such as those used in language classrooms, or medium-stakes tests/small-scale high-stakes tests, such as those used in Intensive English Programs. Again, as with textbook development, corpora has not been used to its full potential.

Csomay (2005) conducted a study that used factor analysis to define the dimensions of university language. She found three dimensions: (a) contextual orientation versus conceptual orientation; (b) personalized framing; (c) interactive dialogue versus monologic (teacher) dialogue. Personalized framing consisted of the use of mental verbs, elements that indicate stance (verbs that take THAT complement clauses), and elements of personal narrative. She next looked at whether university class sessions more closely resembled spoken or written discourse (Csomay, 2006). Two of the features that she examined were the use of first and second person pronouns and demonstrative pronouns. She found that class sessions were closer to conversations than to written discourse.

Csomay (2007) also looked at how speaker, level of instruction, and discipline influenced the production of the features from the three dimensions of university language that she proposed. She found that discipline does influence features in the dimension she labeled personalized framing, with education, humanities, and social sciences lecturers more likely to express their thoughts than engineering, business, and natural sciences lecturers.

Biber (2006) also examined university language. He found that modal verbs most common grammatical way to mark stance in both written and spoken university registers. He also looked at stance verbs + *to* clauses and stance verbs + *that* clauses and found that both were more common in spoken university registers than in written registers.

Research Question

This study looked at listening test texts from an Intensive English Program (IEP). It was an exploratory study that looks at a range of features of the corpus. The features selected for examination were all based on prior research into spoken academic English and include: (a) subject pronouns, (b) modal verbs, (c) stance verbs of desire (*want, like, love*) + “to” clauses, and

(d) “that” clauses (i.e., “that” relative clauses, noun complement clauses, adjective complement clauses, and verb complement clauses). The research question driving this study was: Are there differences across the registers in the IEP corpus?

Methods

Corpus Description

The corpus came from placement and exit listening tests at a North American IEP. The first step in developing the corpus was to obtain permission from the IEP to use the tests. Once permission was granted, the tests were obtained as MP3 files. The files were first submitted to an automatic transcription program. They were then manually checked against the audio files and transcription errors were corrected. Each file was coded for certain information, including word count and the semester it was administered. Table 1 presents basic information about the corpus.

Table 1

Corpus Information

	Face-to-face interactions	Media interactions				Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
Number of texts	13	1	1	4	2	13	1	2
Word count	6229	245	823	2435	1287	9789	720	565
Number of texts by group	13		8				16	
Word count by group	6229		4790				11074	
Percent of corpus	28%		22%				50%	

Note. CON = conversation, ADV = advertisement, RAD = radio show, NEW = news report, INT = interview, LEC = lecture, PRE = presentation, TLK = talk; Total words in entire corpus: 22,093

The corpus included 37 listening test texts (22,093 words) in total. While the texts all existed on the spectrum between monologic and dialogic, no text was purely monologic; all texts included at least some interaction.

The registers included conversation (13), news (4), radio shows (1), advertisements (1), interviews (2), lectures (13), talks (2), and presentations (1). The decision was made by the researcher to combine the different registers into three main groups based on the type of interaction: face-to-face interactions (conversations), media interactions (news, radio shows, advertisements, and interviews), and interactions between one person and a group of people (lectures, talks, and presentations).

Analysis

To analyze the corpus, the researcher wrote a series of programs using Perl. Precision and recall were checked for each of the programs by comparing the program results against one of the text files. Changes were made to the programs to ensure an acceptable rate of precision and recall.

The first program generated word lists for each of the registers. The second program generated word counts for each of the texts and for the three groups. The next program calculated the number of modal verbs in each text. After that, a program was written to count the first, second, and third person pronouns in each of the texts. The fourth program searched for stance verbs of desire followed by a “to” clause. The fifth program produced a KWIC files of each occurrence of “that.” Each occurrence was hand-coded by the researcher, and an additional program was written to read in the codes and produce an output file for SPSS.

Corpus data were analyzed two ways. The first was by looking at each register individually. Next, registers were combined into the three groups, and data were analyzed by

group. A series of one-way ANOVAs were performed to look for statistically significant differences among registers and groups. Both sets of results are presented below.

Results

The purpose of this study was to explore select features of placement and exit tests used in a North American IEP. The goals were to compare the results to previous research and also to compare results of the three groups against each other. Tables 2 – 8 present the results of this study in a text-linguistic tradition, allowing comparisons both within and across registers (Biber, 2012).

Table 2 shows the results of the word frequency counts. The most frequent ten words for each register were, for the most part, function words, and they appear to be similar across the registers.

Table 2

Word Frequency: Ten Most Frequently Used Words by Register

CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
the	you	a	the	the	the	the	the
to	this	of	a	to	of	and	of
you	your	it	to	of	to	to	to
I	to	the	of	that	and	in	and
and	pill	and	in	is	a	of	will
a	and	in	and	I	that	is	you
that	a	you	is	you	in	a	your
of	I	it's	that	and	is	water	that
in	the	that	you	a	you	on	we
it	don't	to	for	what	I	it	buy

Tables 3-5 show total and normed counts for first, second, and third person pronouns, respectively. Conversations appear to use more first and second person pronouns (normed counts are 13.65 and 13.24, respectively).

Table 3

First Person Subject Pronouns (I, we)

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
1 st Person Pronouns	170	4	21	26	33	215	14	8
Total	170		84			237		
Normed per 500 words	13.65		8.77			10.7		

Table 4

Second Person Subject Pronoun (you)

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
2 nd Person Pronouns	165	9	18	26	21	134	5	9
Total	165		74			148		
Normed per 500 words	13.24		7.72			6.68		

Use of third person pronouns appeared to be equal across the registers. A series of one-way ANOVAs, however, showed no statistically significant difference for first person pronouns between registers ($F = 1.61, p = .18, df$ between groups = 6, df within groups = 31) or groups ($F = 1.33, p = .28, df$ between groups = 2, df within groups = 35). There was no statistically significant difference for second person pronouns among registers ($F = .80, p = .58, df$ between groups = 6, df within groups = 31) or groups ($F = 2.12, p = .14, df$ between groups = 2, df within

groups = 35); and no statistically significant difference for third person pronouns among registers ($F = 1.34, p = .27, df$ between groups = 6, df within groups = 31) or groups ($F = .43, p = .66, df$ between groups = 2, df within groups = 35).

Table 5

Third person subject pronouns (he, she, they)

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
Third Person Pronouns	423	11	65	154	97	621	72	43
Total	423		327				736	
Normed per 500 words	33.95		34.13				33.23	

Table 6 presents information about frequency and normed counts of modals. Face-to-face conversations contained 4.66 modals/500 words, media interaction contained 4.70 modals/500 words, and interactions between one person and a group of people contained 5.91 modals/500 words. A one-way ANOVA showed no statistically significant differences among registers ($F = 1.13, p = .37, df$ between groups = 6, df within groups = 31) or groups ($F = 1.27, p = .29, df$ between groups = 2, df within groups = 35).

Table 6

Modals (can, could, may, might, should, will, could, shall)

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
Modals	58	3	6	28	8	110	1	20
Total	58			45			131	
Normed per 500 words	4.66			4.70			5.91	

Table 7 displays frequency and normed counts for stance verbs of desire plus “to” clauses. The first two groups contained roughly the same number of these per each 500 words (0.96 and 0.42, respectively). The third group, interactions between one person and a group of people, had 1.35/500 words.

Table 7

Stance verbs of desire (want, love, like) + “to” clause

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
Stance Verb	12	0	0	2	2	26	1	3
Total	12			4			30	
Normed per 500 words	0.96			0.42			1.35	

Results of the one-way ANOVA to look for differences among register were not statistically significant ($F = .731$, $p = .63$, df between groups = 6, df within groups = 31). Results

of the one-way ANOVA to look for differences among groups were also not statistically significant ($F = 2.08$, $p = .14$, df between groups = 2, df within groups = 35).

Tables 8 and 9 display the results of the examination of the different “that” clauses. Table 8 shows results by register. Both interviews and lectures have higher normed counts for relative clauses (2.72 and 2.96, respectively) and for verb complement clauses (3.50 and 2.86, respectively).

Table 8

“that” clause by register

	Face-to-face interactions		Media interactions			Interactions between one person and a group of people		
	CON	ADV	RAD	NEW	INT	LEC	PRE	TLK
NCC	8 0.64	1 0.10	3 1.82	9 1.85	5 1.94	32 1.63	0 0.00	0 0.00
RC	17 1.36	1 0.10	0 0.00	9 1.85	7 2.72	58 2.96	3 2.08	2 1.77
ACC	4 0.32	0 0.00	0 0.00	0 0.00	1 0.39	1 0.05	0 0.00	1 0.88
VCC	21 1.69	0 0.00	3 1.82	10 2.05	9 3.50	56 2.86	2 1.39	2 1.77

Note. Top number is frequency, bottom number is normed to 500 words. Register: CON = conversation, ADV = advertisement, RAD = radio show, NEW = news report, INT = interview, LEC = lecture, PRE = presentation, TLK = talk. “that” clauses: NCC = noun complement clause, RC = “that” relative clause, ACC = adjective complement clause, VCC = verb complement clause

Table 9 displays results by group. Overall, adjective complement clauses were the least-frequently-occurring. The most frequently-occurring clause types varied by group. Noun complement clauses and relative clauses occurred most frequently in media interaction. Verb

complement clauses occurred most frequently in the third group, interaction between one person and a group of people.

Table 9

“that” Clauses by Group

	Face-to-face interactions	Media interactions	Interactions between one person and a group of people
NCC	8 0.64	18 1.88	32 1.44
RC	17 1.36	63 6.58	17 0.77
ACC	4 0.32	2 0.21	1 0.05
VCC	21 1.69	22 2.30	71 3.21

Note. NCC = noun complement clause, RC = “that” relative clause, ACC = adjective complement clause, VCC = verb complement clause

A series of one-way ANOVAs were performed to see if use of the four different clauses types varied significantly by register or group. Table 10 presents the results of the analyses.

There were no statistically significant findings from the tests.

Table 10

One-way ANOVAs

	Register		Group	
	<i>F</i>	Sig.	<i>F</i>	Sig.
“that” noun complement clause	.86	.54	.97	.39
“that” relative clause	1.26	.30	.79	.46
“that” clause controlled by an adjective	.22	.97	.42	.66
“that” clause controlled by a verb	.23	.96	.14	.87

Note. *df* for register (6, 29); *df* for group (2, 33)

Relevance to PIE and Second Language Learning

The primary importance of this study for the PIE at NAU is that it created a corpus of the placement and exit listening tests which can be used for research. The results of this study offer preliminary information on some features of the NAU PIE listening tests. Although this study did not find any significant differences among registers or texts, it was not exhaustive. It does indicate, however, the major differences may exist in areas other than grammar. This alone offers directions for future research, which could look at features of spoken language such as sandhi variation, hesitations and false starts, and rate of speech. Also, rather than analyzing texts by register, they could be grouped as they appear on tests. This might be a way to check the equivalency of different test forms. Finally, texts can be compared against test-taker scores to see how each text performs.

Future directions for research, however, rest upon finding differences among the texts. The first question is, what are the differences? The next two questions diverge. The first has implications for test development: how do these differences link to test-taker performance. In other words, what is the relationship between text features and test scores? The other question has implications for the construct validity of the test: How do the features of the texts reflect (or fail to reflect) the features found in spoken university language (including conversations, service encounters, and lectures)? One final question remains: Once IEP students are exited from the program and enter university, how well-prepared are they to handle the demands of spoken university language?

References

- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5, 97-116. doi:10.1016/j.jeap.2006.05.001
- Burton, G. (2012). Corpora and coursebooks: Destined to be strangers forever? *Corpora*, 7(1), 91-108. doi: 10.3366/cor.2012.0019
- Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52(1), 43-56). doi: 10.1093/elt/52.1.43
- Csomay, E. (2005). Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education*, 15, 243-274. doi:10.1016/j.linged.2005.03.001
- Csomay, E. (2006). Academic talk in American university classrooms: Crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes*, 5, 117 – 135. doi:10.1016/j.jeap.2006.02.001
- Csomay, E. (2007). A corpus-based look at linguistic variation in classroom interaction: Teacher talk versus student talk in American University classes. *Journal of English for Academic Purposes*, 6, 336 – 355. doi: 10.1016/j.jeap.2007.09.004
- McNamara, T., Hill, K., May, L. (2002). Discourse and assessment. *Annual Review of Applied Linguistics*, 22, 221-242. doi: 10.1017/s0267190502000120