ABSTRACT

Title of Dissertation:	SPEECH MODIFICATION TO NON-NATIVE SPEAKERS AND CONTENT DILUTION: IMPLICATIONS FOR ENGLISH AS A MEDIUM OF INSTRUCTION (EMI)
	Assma Mohammad Al Thowaini, Doctor of Philosophy, 2018
Dissertation directed by:	Dr. Michael H. Long, Second Language Acquisition

With the rapid growth of language education programs, such as Content-and-Language-Integrated Learning (CLIL) and English as a medium of instruction (EMI), research on input modification shifted perspectives. The current study investigates L2 input modification by comparing the speech of non-native speakers to that of native speakers towards low-proficiency learners of English using quantitative methods. Furthermore, the study explores the effects of these modifications on learners' content comprehension and the possible content dilution (e.g., loss of essential information) triggered by linguistic simplification. In this experiment, two types of participants were recruited: speakers and listeners. Twenty native and advanced non-native speakers of English participated (ten of each). The speaker participants were divided into two subgroups: those with language teaching experience and those without. For the listeners, three groups were recruited: 20 native speaker controls, 20 high-

proficiency, and 20 low-proficiency English learners (listeners). Each speaker narrated stories to three assigned listeners (one from each condition) in one-on-one sessions. Each session included an introduction, two warm-up stories, and three main stories. Speech was audio-recorded to examine the types of modification employed with high- and low-proficiency listeners, as opposed to native listener controls, and the effects of those modifications on story content and listener comprehension. After each story, the listeners took a content comprehension assessment.

The transcripts were coded for lexical complexity (diversity and sophistication), syntactic complexity, and content dilution. The results showed a significant difference between native and non-native speakers in their speech to the three listener conditions in terms of lexical sophistication and syntactic complexity, as well as a significant difference between speakers with language teaching experience and speakers without in terms of lexical diversity. Furthermore, all speaker conditions exhibited significant linguistic modification (lexical diversity, lexical sophistication, and syntactic complexity) in their speech towards low- and high-proficiency listeners compared to their speech towards the native controls. In addition, only native speakers showed significant content dilution (measured by the count of mentioned information bits) in their speech towards high- and low-proficiency listeners. Finally, the high- and low-proficiency listeners' scores on the content comprehension assessment were significantly lower than the scores of the native controls.

SPEECH MODIFICATION TO NON-NATIVE SPEAKERS AND CONTENT
DILUTION: IMPLICATIONS FOR ENGLISH AS A MEDIUM OF
INSTRUCTION (EMI)

by

Assma Mohammad Al Thowaini

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
 Professor Michael H. Long, Chair
 Professor Robert M. DeKeyser
 Professor Steven J. Ross
 Professor Jeff MacSwan
 Professor Min Wang

Dedication

To my parents ..

إلى من غمروني بحبهم ورعايتهم طول حياتي ..

إلى من أكرموني ودعموني بلا حدود ..

إلى من علموني معنى الصبر والكفاح والتضحية ..

إلى والدي ووالدتي ..

Acknowledgment

<div dir="rtl">

بسم الله الرحمن الرحيم

اللهم لك الحمد كما ينبغي لجلال وجهك وعظيم سلطانك. اللهم لك الحمد حتى ترضى ولك الحمد إذا

رضيت ولك الحمد بعد الرضا، الحمد لله الذي بنعمته تتم الصالحات.

وصلى الله على نبينا محمد وعلى آله وصحبه أجمعين.

</div>

First and foremost, all praise and gratitude are due to Allah, God, for only through His grace and blessings that I was able to complete my doctoral studies.

Secondly, I am grateful to the committee members for their time and feedback throughout the dissertation process. Special thanks to Professor Michael Long, my dissertation chair, for directing and supervising the various stages of the project. I am also grateful to Professor Steve Ross whose knowledge and expertise guided the analytical and technical stage of the study. Additionally, my deepest thanks go to Professor Robert DeKeyser for his insightful comments. I am equally grateful to both Professor Jeff MacSwan and Professor Min Wang for their valuable feedback.

Furthermore, I want to acknowledge the encouragement, mentor-ship, and friendship of various individuals whom I met during my journey at UMD. Particularly, Professor Kira Gor, Professor Nan Jiang, Professor Peter Glanville, Professor Lauretta Clough, Josiland Chambers, Katie Kiraly, Man Li, Kyoko Kobayashi, Tom Wagener, Suzanne Freynik, Alia Lancaster, Yuichi Suzuki, and Payman Vafaee. Thank you very much.

My heartfelt thanks go to my colleagues and friends at the Pennsylvania State

Table of Contents

List of Tables

List of Figures

List of Excerpts

Introduction

Language and language education have been important for long periods of time. In the 16th and 17th centuries, numerous Europeans traveling in the Ottoman and Safavid empires in search of trade, knowledge, political interests, or religious pilgrimage had to learn additional languages (i.e., Arabic, Turkish, and Persian) or employ interpreters in order to communicate (Brentjes, 2010). For instance, the German traveler Leonhard Rauwolf (1535-1596) was not proficient in Arabic or Turkish and depended on merchants and interpreters to help him learn and understand the local languages. On the other hand, the astronomer Zayn al-Din Lari met with Pietro della Valle (1586-1652) in the province of Fars and learned sufficient Latin from Della Valle to understand and interpret the scientific books Lari's Italian friend sent him (Brentjes, 2010).

While the methods have changed dramatically, the objectives of language education remain similar today with more focus on communicative goals. Education policies worldwide have adopted different political and educational approaches to align with modernization, and language education has become a key part of many countries' formal education systems. Globalization has made it necessary for individuals in numerous workforces (e.g., medicine, journalism, tourism, business, and international relations) to use and communicate through common languages. Due to the military, economic, and political power of English-speaking countries since the 18th century, English has prevailed as the global *lingua franca*, and many countries have implemented English as a foreign language (EFL) education in the public, private, and post-secondary education sectors. Although many English-teaching approaches exist, one of the main methods is to incorporate English as the medium of instruction (EMI) for content learning.

This study explores the influence of content-and-language-integrated instruction on content learning. Specifically, it investigates, in a laboratory setting, how native and non-native English-speakers use linguistic modification in narrating stories to listeners of varied English proficiency, and the extent to which such modification results in content dilution. This chapter situates EMI in the SLA literature. Considering the wide range of research conducted in different contexts, the scope of the review is limited to studies that discuss EMI in the context of Saudi Arabia. The chapter concludes with an in-depth review of input modification literature over the past four decades and discusses linguistic complexity measures employed by researchers.

Chapter 1: Literature Review

1.1 L2 as a Medium of Instruction

After the success of French immersion programs in Canada, as reported by many researchers over the past 40 years (e.g., Genesee, 1987; Johnson & Swain, 1997; Lazaruk, 2007; Swain & Lapkin, 1982, 1986), models integrating language and content learning started to increase rapidly in the mid-1990s (Dalton-Puffer, 2007; Rodríguez, 2012). Marsh (2002) described these initiatives as "any dual-focused educational context in which an additional language, thus not usually the first language of the learners involved, is used as a medium in the teaching and learning of non-language content" (p. 15). He added that "if there is no dual focus on language and non-language content within a lesson or course, then it does not qualify as a form of CLIL [Content-and-Language-Integrated Learning]" (Marsh, 2002, p. 17). Research on this educational approach, however, has applied the theory variously, with different contexts, objectives, and definitions of what constitutes L2 as a medium of instruction. In particular, contexts explored in studies of CLIL differed from those in studies of EMI. CLIL is a broader term that may include EMI and Integrating Content and Language in Higher Education (ICL or ICLHE) to raise "referential distinctions which can be traced to different conceptualisations of the term" (Smit & Dafouz, 2012, p. 6) and to differentiate it from common CLIL settings. Other conceptualizations or models include Content-Based Language Teaching (CBLT), bilingual education, two-way immersion, and sheltered instruction (for reviews see: Cenoz, 2014; Long, Al Thowaini, Al Thowaini, Lee, & Vafaee, 2018; Lyster & Ballinger, 2011; Mauranen, 2012; Papaja, 2014; Tarnopolsky, 2013).

1.1.1 Contexts of content-and-language-integrated initiatives

Studies of both EMI and CLIL focus on moderating variables ranging from the context to the students' and instructors' L2 proficiency (Lo & Lo, 2014; Papaja, 2014), but both CLIL and EMI have one common characteristic: students' L2 proficiency may be inadequate to learn content material through that language. Although the research from both contexts can overlap, the work is fundamentally different, and studies conducted with each model may have mixed results. The main difference lies in the contexts in which researchers interpret their findings, which are distinguished by six features: (a) theoretical framework and origin (Dearden, 2014), (b) L2 medium, (c) context and setting, (d) type of content, (e) type of language instruction, and (f) instructors' background and language proficiency.

CLIL originally started in Europe and was intended to develop both the language and the content being taught (Aguilar & Munoz, 2014; Coyle, Hood & Marsh, 2010; Dalton-Puffer, 2007, 2011; Lasagabaster, 2011; Marsh & Wolff, 2007). CLIL contexts do not have one specific language used as the medium of instruction. While some studies use English (e.g., Lasagabaster & Sierra, 2009; Mehisto, Marsh & Frigols, 2008; Navés & Victori, 2010), others use Basque (e.g., Cenoz, 2008, 2014), Catalan (e.g. Arnau, 2000), Swedish (e.g., Södergård, 2008), or many other languages. In addition, CLIL is implemented in the early stages of education, mainly in elementary and secondary schools (Papaja, 2014), and the instructors, as well as the students, typically have low to intermediate proficiency in the language used. Instructors have generally been trained to teach the subject matter, but not in a different language (Long et al., 2018). Therefore, the content may be diluted through instructors' inexperience teaching in a non-native language to relatively low-proficiency learners.

4

EMI does not have a theoretical framework for its implementation (Chapple, 2015; Dearden, 2014). While CLIL was established on the basis of teaching both language and content (Marsh, 2002), EMI developed instead from a practical need to be able to use English, the global *lingua franca*. As English's worldwide hegemony became clearer, EMI started simultaneously in different regions with the aim of teaching content through English (Byun et al., 2011; Dearden, 2014; Macaro, Akincioglu & Dearden, 2016). Chapple (2015) described EMI as "the single most significant current trend in internationalizing higher education" (p. 1). The rapid pace of EMI implementation has been described by Oxford as "an unstoppable train which has already left the station" (Macaro, 2017, p. 2) and as a potential "pandemic" (Phillipson, 2009). Galloway, Kriukow, and Numajiri (2017) listed a number of possible driving forces behind implementing EMI, including gaining access to innovative knowledge, increasing a university's global competitiveness, increasing student and instructor mobility, "enhancing the employability of graduates' competencies, and improving English proficiency" (p. 1). Whereas the abovementioned motives behind EMI seem sensible, it is still unclear whether EMI has any effect on English language or content learning.

Dearden (2014) conducted an exploratory 55-country survey of the EMI phenomenon to map the size and shape of EMI education. She defined EMI as "the use of the English language to teach academic subjects in countries or jurisdictions where the first language (L1) of the majority of the population is not English" (p. 4). This definition is broad, and it does not fully describe the context of EMI or its distinctions from CLIL (Europe) or immersion (Canada) and bilingual programs (United States). Rather, EMI should be defined as teaching scientific subjects (e.g., subjects in science, mathematics, engineering, and medicine) through English in

primarily higher or tertiary education where the context is not English-speaking and the majority of students' L1 is not English. Additionally, in EMI, there is no explicit language learning, and the instructors are either native English speakers or non-native speakers who have received their degrees from English-speaking countries and/or have worked in an EMI setting (i.e., they are familiar with the content in English). Finally, most students have already studied English as a foreign language during their primary and secondary education, and they often go through a transitional year to prepare for EMI university courses.

Clearly, EMI and CLIL contexts are not identical to one another (see Table 1). Furthermore, there are differences found within each type. This variation in EMI is similar to differences within CLIL contexts. Within European education, for instance, CLIL contexts differ (Ruiz de Zarobe & Lasagabaster, 2010) on the basis of students' ages, types and number of classes taught, and selection/admission method. Therefore, it is important for researchers to describe in detail the context in which a study is conducted and which type of content-and-language-integrated implementation is used; taking account of such information is essential for accurate experimental designs, appropriate measures, precise procedures to yield comparable results, and clearly delineated generalizations of the results. While the current study was conducted in a laboratory, not an EMI context, the above-mentioned program descriptions were thoroughly explored to design an experiment that closely simulates these contexts (instructors' background, students' proficiency levels, etc.).

Feature	EMI	CLIL	Immersion
Aims	Not clear	Clear	Clear
L2 medium	English	English, Basque, Catalan, Swedish or other languages	Originally French, but now includes other languages
Context	Middle East, Eastern and Western Asia, and Europe	Mainly Europe	Mainly Canada, but elsewhere, as well
Setting	Higher education (scientific, engineering, and medical fields)	Mostly elementary to primary education	Consistent from early to late stages of learning
Type of content taught	Degree-related: medicine, engineering, mathematics, and science courses	Social studies, arts, science, mathematics, health, and electives	L2 curriculum parallel to L1 curriculum
Type of language instruction	No concurrent language instruction. The students have taken EFL classes during their primary education and through their preparatory year	Instruction varies. Separate or concurrent language instruction	Concurrent language instruction
Instructors' L2 proficiency	Ranges from intermediate to advanced	Usually intermediate	Mostly advanced
Instructor background	MA and Ph.D. holders – many sought their degrees from English-speaking countries	BA holders – many sought their degree from local universities or teacher colleges	Usually trained bilingual instructors
Students' L2 proficiency	Heterogeneous proficiency	Heterogeneous proficiency	Mostly homogeneous proficiency

Table 1: Differences between EMI, CLIL, and Immersion settings

1.1.2 Lack of empirical support for EMI

SLA and applied linguistics scholars continue to debate whether EMI qualifies

as a method of language teaching. Researchers and educators often generalize the

success of immersion programs to CLIL and EMI contexts. Immersion programs in

Canada have mainly succeeded because of numerous distinctive features (as indicated

in Table 1), including trained bilingual instructors, early L2 exposure, students'

homogenous L2 proficiency, and an L2 curriculum that parallels the L1 curriculum

(Cummins, 1998; Johnson & Swain, 1997; Long et al., 2018). Met (1998) suggested a

continuum, later adapted by Lyster and Ballinger (2011), with the aims of language

learning (language-driven) at one end, and of content learning (content-driven) at the

other (see Figure 1). Certainly, EMI is placed towards the content-driven end, even though students are not partially immersed, and it is distant from the language-driven end, as there is no explicit language instruction whatsoever. Based on this continuum and the above-mentioned distinctive features, EMI is very different from immersion programs, whose success has been well documented.

Content-driven			Language-driven		
Total immersion	Partial immersion	Content courses	Content courses + language classes	Language classes with thematic units	Language classes with content used for language practice
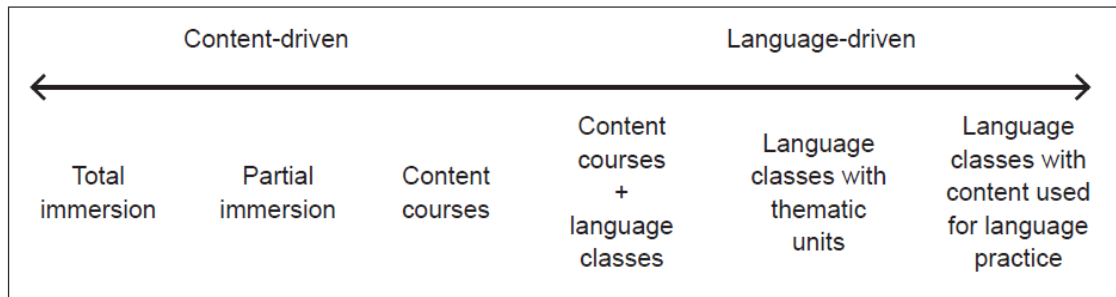
* From Lyster & Ballinger (2011, p. 280)
Figure 1: Range of content-and-language-integrated settings

Despite the differences in contexts, content-and-language-integrated initiatives are commonly found to result in greater language gains than traditional language teaching. While no research has explored language gains in EMI contexts, many CLIL studies have been conducted to support such claims. However, these studies have suffered from major methodological problems and sample selection bias (for critical reviews: see Bruton, 2011, 2013, 2015). For instance, Aguilar and Munoz (2014) investigated the English-language outcomes of a 60-hour engineering course. In a pretest/posttest design, they compared high- and low-proficiency learners' listening and grammar gains. While the results for both high- and low-proficiency learners showed no improvements in grammar comprehension, low-proficiency learners' scores yielded a significant gain in listening comprehension. The researchers suggested that low-proficiency learners benefit more from a CLIL setting. Nonetheless, it was not clear what types of language lessons, or lack thereof, were used in this setting and whether instructors modified their speech. In addition, as with

many other CLIL studies, the researchers did not assess the learners' gains in comprehension of the subject matter.

Dallinger, Jonkmann, Hollm, and Fiege (2016) considered these and other methodological constraints and conducted a longitudinal study in Germany controlling for selection effects (i.e., student, classroom, and teacher characteristics). With 1,806 German CLIL and non-CLIL eighth-graders in English and history, the researchers investigated a wide range of skills and variables, including English listening comprehension, general English proficiency, history (content) learning, cognitive ability, and motivation. Confirming CLIL selection effects, they found that CLIL students had better prior achievement, higher motivation, higher cognitive abilities, higher socioeconomic status, and even better instructional quality in history. Controlling for these variables, the results showed significantly greater listening comprehension achieved by CLIL students. However, the results for general English proficiency and content learning showed comparable development between CLIL and non-CLIL classrooms. The researchers noted that CLIL classrooms invested substantially more time (50% more) in content learning and yet only achieved comparable learning outcomes, suggesting a major limitation of CLIL for subject matter learning.

Irrespective of whether scholars consider EMI a method of language teaching, educational institutions that use EMI do not provide empirical support for the initiative, "which means the trend is developing in a fairly 'organic' manner" (Madhavan & McDonald, 2014, p. 2). Just as with CLIL, little empirical research has been conducted to assess or validate EMI by showing evidence of language and content learning. Also, studies of the efficiency of EMI for improving English language competence in comparison to EFL instruction are scarce (Byun et al., 2011;

Hu & Lei, 2014). That is, there is not enough information regarding the consequences of using EMI in higher education for "teaching, learning, assessing, and teacher professional development" (Dearden, 2014, p. 4). Moreover, policymakers show little concern about students' inadequate English competence, which results in major shortcomings in content comprehension (Byun et al., 2011; Macaro, 2014). Overall, EMI implementation is generally hasty and unsystematic and lacks evidence of its "effectiveness as a new pedagogical method for promoting English (as opposed to EFL) [and it] disregards the danger that students will not understand the content that is being presented to them" (Macaro et al., 2016, p. 52).

Although little quantitative research has been conducted on EMI, many studies within the last decade have qualitatively explored the initiative in different contexts worldwide. These studies have taken one of two main approaches. The majority explored teachers' and/or students' perceptions and attitudes towards the use of EMI, including the challenges they encountered and the strategies they developed to cope with its implementation (e.g., in *China*: Hu & Lei, 2014; *Denmark*: Werther, Denver, Jensen & Mees, 2014; *Egypt*: Sabbour, Dewedar, & Kandil, 2010; *Indonesia:* Floris, 2014; *Iran*: Zare-ee & Gholami, 2013; *Korea*: Kym & Kym, 2014; *Saudi Arabia*: Al-Kahtany, Faruk, & Al Zumor, 2016; Shamim, Abduelhalim, & Hamid, 2016; *Spain*: Morell et al., 2014; *Taiwan*: Chang, 2010; Yeh, 2012; *Turkey*: Kilickaya, 2008; Macaro et al., 2016; *Ukraine*: Goodman, 2014; *United Arab Emirates*: Belhiah & Elhami, 2015; *Vietnam:* Manh, 2012; Vu & Burns, 2014). The second approach has investigated the challenges of language policy and EMI implementation globally and locally (e.g., Byun et al., 2011; Coleman, 2006; Dearden, 2014; Doiz, Lasagabaster, & Sierra, 2013; Galloway et al., 2017; Hu, Li, & Lei, 2014; Li, Leung, & Kember, 2001; Shamim, 2011; Shohamy, 2013). A review of these studies raises three

overarching concerns about EMI: (1) students' inadequate English proficiency, (2) lack of a theoretical framework, and (3) lack of systematic implementation. In an attempt to identify and describe one of many EMI contexts, these issues are discussed in the following section based on studies undertaken in the context of Saudi higher education. The context of Saudi Arabia was selected to reflect the laboratory study's main sample: native speakers of Arabic as non-native English speakers (instructors) and listeners (students).

1.2 The status of English in Saudi Arabia

EFL is commonly taught in elementary and primary education in countries under the Gulf Cooperation Council (GCC), such as Kuwait, Saudi Arabia, United Arab Emirates (UAE), and Qatar. Interestingly, in Saudi Arabia, the government for various reasons took the initiative to include English in public education shortly after its establishment in 1932 (Al-Seghayer, 2005). The most important reason at that time was that Saudi Arabia had become a prominent supplier in the oil industry and exported oil to many countries worldwide where English was a common language in trading. Thus, the government realized the importance of preparing its citizens to communicate with the English-speaking populations and to be able to establish cross-cultural understanding by working, in some cases, with Westerners employed by the Arabian-American Oil Company (ARAMCO).

In addition, the government launched large-scale developmental programs in the early 1970s to create an infrastructure that included transportation, water, electricity, telecommunications, health, social welfare, and education. Such programs required recruiting manpower that used English for communication. Therefore, educational policymakers recognized the significance of introducing English into the curricula "to prepare personnel who could fill positions requiring knowledge of English" (Al-Seghayer, 2005, p. 134). Furthermore, every year, approximately two

million non-Arab Muslims from all over the world come to Saudi Arabia to perform

pilgrimage (i.e., *hajj*, fulfilled only during the last month of the Islamic lunar

calendar, and *umra*, executed any other time of the year), so it is essential for Saudis

to learn English as a *lingua franca,* in order to communicate and interact with these

visitors (Al-Seghayer, 2005). For these reasons, English is taught as a foreign

language in public schools throughout Saudi Arabia. The Saudi curriculum includes

English in two ways: (a) English as a foreign language (EFL) and (b) English as a

medium of instruction (EMI). The following sections discuss each implementation in

detail.

1.2.1 English as a Foreign Language

Similar to those in other regions in the world, the Saudi public school system

comprises twelve years of education. This includes elementary/primary school (six

years), middle/intermediate school (three years), and high/secondary school (three

years). Students receive the same curriculum up to tenth grade, with EFL as a

compulsory subject from fourth grade until tenth. Once students reach tenth grade in

high school, they enter one of two academic streams: (a) scientific or (b) literary

education. In order to be admitted to a stream, students must have achieved a specific

level of academic performance. Depending on the stream students enter, they may

continue to take English subjects in grades 11 and 12.[1]

Presently, English is the only foreign language included in Saudi public and

private schools, universities,[2] and governmental, as well as industrial, institutions.

English is required as either an elective or as part of a major field of study—e.g.,

English literature or translation—in universities. Even non-English fields, such as

---

[1] For further information, see http://www.moe.gov.sa/ar/Pages/default.aspx
[2] In Language and Translation colleges in various universities, students can seek degrees in other languages, such as French.

sharia (Islamic law), *da'wah* (sharing and teaching Islamic values), geography, and Arabic, require preparatory courses in English. Even though most Saudis have studied English in public schools for at least seven years, academic institutions for higher education require a minimum of one introductory English course. This emphasis on English reflects the current prevalence of job requirements that applicants have a certain level of English proficiency.

English curricula and textbooks are nationally standardized in Saudi public schools. Most Saudi schools are segregated by gender, but both boys and girls receive similar educational programs. English is taught four times a week, and each period lasts 45 minutes. The curriculum is constructed and the textbooks are published by the Ministry of Education (MoE), which employs curriculum development experts. Thus, learning EFL in Saudi Arabia is based mainly on the MoE textbooks. Many English teachers at these public schools are native Arabic-speakers who have graduated from local teacher colleges, where they attended four-year programs for English teachers. They are trained to teach the language using the Audio-Lingual Method and/or Grammar Translation (Al-Seghayer, 2005; Al-Mohanna, 2010). Instruction is primarily based on the MoE textbooks, in teacher-centered classrooms (Al-Mohanna, 2010). Teachers lecture, "drill," and repeat and write grammar rules and examples on the blackboard. They also depend on Arabic to explain (i.e., translate) some of the English rules to students. Furthermore, there is little emphasis on collaborative work or individual creative projects. In order to assess students, teachers give monthly quizzes, midterms, and final exams for each marking period.

Outside of the EFL classroom, students have little opportunity to practice English. Arabic is the single national language of Saudi Arabia and is used in most public, workforce, and educational domains (e.g., shopping centers, banks,

government offices, hospitals, and schools), with the exception of a few international schools and universities (Al-Nofaie, 2010; Shamim, Abduelhalim & Hamid, 2016).

Arabic is also the medium of instruction throughout Saudi public education, except for EFL classes. Because of the limited opportunity for students to practice English and other constraints (i.e., types of pedagogical implications and students' learning styles), many students graduate from high school with English proficiency that does not meet the minimum standards for admission to higher education or the workforce (as reflected in the sample of low-proficiency participants for the current study). These institutions and employers require a minimum ability in English for various reasons, including university rankings,[3] global mobility, and internationalization. Depending on what degrees students seek, their English proficiency could determine their performance in their chosen fields. This is primarily true in the many scientific fields that require instructors to teach content using English as a medium of instruction.

1.2.2 English as a medium of instruction (EMI)
English proficiency remains essential for higher education, as well. Admission to certain university programs requires that students meet a number of criteria. These comprise, but are not limited to, high school GPA, General Aptitude Test (GAT, known as *quduraat* in Saudi Arabic) score, Scholastic Achievement Admission Test (SAAT, or *taHseel*)[4] score, and Standardized Test of English Proficiency (STEP) score. Contingent upon the degree students seek, a compound percentage of these scores is considered. For instance, while a degree in Islamic studies requires a percentage combination score (weight) of GAT (40%) and high school GPA (60%), a

---

[3] English would help garner a university higher rankings by increasing publications in high-quality and well-known journals, most of which are published in English.
[4] All required tests are developed and monitored by National Center for Assessment in Higher Education. See http://www.qiyas.org for further information.

degree in a scientific field requires high school GPA (30%), GAT scores (30%), and SAAT scores (40%; Admission Standards, 2016). Once students are admitted, all degrees entail a one-year preparatory program that focuses primarily on English, as well as some basic math and science courses. Some studies have investigated the quality of these preparatory programs and whether their learning outcomes align with their objectives, such as preparing students for EMI courses (e.g., Aburizaizah, 2015; Kabouha & Elyas, 2015; McMullen, 2014; Shamim et al., 2016).

In the most recent of these, Shamim et al. (2016) closely examined the EMI setting in the transition year,[5] known as the Preparatory Year Program (PYP). In the PYP, EMI is used for science subjects in all universities in Saudi Arabia. Researchers observed lessons taught by seven instructors and conducted follow-up interviews. All instructors but one were fluent in both English and Arabic. Nineteen students were assigned to three focus groups of six to eight, based on their English proficiency. Researchers expected that students' proficiency levels would determine their perceptions of the English language, and this was confirmed by the results. Whereas students with higher proficiency favored the use of EMI, students with lower English proficiency questioned it because it created a disadvantage for them in achieving admission to their preferred colleges.

Although the study revealed an overall positive perception of using EMI, it also showed some issues with its application. While many instructors and students agreed that learning English was necessary, due to its status as a *lingua franca* in scientific fields, the instructors and students nevertheless had to develop strategies for coping with some of the obstacles EMI presented. As noted by the researchers'

---

[5] The first year of college is called transition or preparatory year. The purpose of this year is to prepare students with the adequate skills to succeed in pursuing their university degrees. The transition year programs include general skills in science and mathematics as well as intensive English courses.

observations, the instructors gave standardized PowerPoint presentations mainly in English, but the classroom lessons, explanation of concepts, and feedback were carried out in Arabic. Because many students were unable to read English textbooks, let alone comprehend unfamiliar content delivered in English, instructors simplified and adjusted texts in several ways, such as developing bilingual word lists, checklists (i.e., they categorized sections from the textbook), and practice exams to help students study. Moreover, students were heavily dependent on Arabic to understand the English content because of their inadequate command of English. Many students could only learn the content by translating all lessons from English to Arabic, forcing instructors, as mentioned above, to accommodate and "tailor" the curriculum through Arabic rather than English. Although unintended, this resulted in allocating the majority of class time to language learning instead of focusing on the quality of content learning.

After passing these preparatory program courses, students in many medical, technical, and scientific fields will continue to use English, as most, if not all, of their major courses are taught with EMI (e.g., King Saud University for Health Science in Riyadh, King Khalid University in Assir, and King Fahad University of Petroleum and Minerals in Dhahran all use EMI). After several years of English education, it is assumed that students are prepared to take EMI courses; however, a recent qualitative study by Al-Kahtany and collaborators (2016) found that undergraduate students still struggle to understand content taught in English, as reflected by their overall academic achievement (e.g., 60% of the students' GPAs were 1-1.99 out of 5). In this study, the researchers explored the attitudes of 702 Saudi students and 162 instructors towards EMI in King Khalid University (located in Assir, a southern province of Saudi Arabia). The study also assessed the influence of linguistic hegemony in the

choice of EMI. All participants were recruited from three major scientific disciplines: a) Science,[6] b) Engineering, and c) Medicine.[7] The data collection was based mainly on a Likert scale questionnaire directed to both instructors and students. The results showed a number of issues that need to be considered. Generally, the instructors' perspectives and values were completely different from those of the students (Al-Kahtany, Faruk, & Al Zumor, 2016). On the one hand, the instructors were largely in favor of EMI and agreed that students were incompetent in English, but they believed that the problem was temporary and that students would overcome this issue when they were provided with an adequate English program. Interestingly, the instructors insisted that English should continue to be the only medium of instruction and did not support using Arabic, despite the difficulty they faced in communicating with their students in English only. The researchers argued that the instructors' favorable attitude towards using English was because they were "blinded by the hegemony of English" (p. 54) in scientific fields.

On the other hand, students showed resistance to and disagreement with using EMI, mainly because they felt that their incompetence in the language led them to receive poor grades in these courses. They perceived English as a barrier rather than a medium, as it made their learning more difficult and hindered their overall achievement. Nevertheless, the students, aware of the importance of English for surviving in this globalized world, were prepared to use English minimally (i.e., they wished to limit the use of English to terminologies only), as long as it did not impede their content learning. Still, many of these students did not see how English would fit in their future, as primarily Arabic is used in many fields of employment in Saudi

---

[6] Includes Natural Sciences, Applied Sciences, and Computer Science
[7] Includes Dentistry, Pharmacology, Medical Sciences, Medicine, and Nursing

Arabia, and they had no plans to study or work abroad in an English-speaking country or to move for employment.

Suliman and Tadros (2011) investigated the coping strategies that nursing students used to deal with EMI at King Saud Bin Abdulaziz University for Health Sciences. In a repeated measures design with the Ways of Coping Questionnaire[8] (WCQ), 78 female student participants completed three questionnaires during three stages of the semester—weeks 3, 9, and 15. The results showed that, at the beginning of the semester, the majority (77%) feared communicating in English mainly because of their lack of fluency in the language. By the end of the semester, confrontive coping[9] and distancing[10] had increased dramatically, suggesting frustration, anxiety, and attempts to avoid thinking about the challenges students were facing with EMI. Finally, the researchers found that participants had begun to employ coping strategies of memorizing, translating, and copying texts without understanding them, a type of learning called *rote learning*. That is, these students had been trained to learn by storing information, such that they were not able to retrieve it in a selective or supportive manner (Dudley-Evans & Swales, 1980). This poses a problem, as nurses or other medical technicians need to be proficient not only in English, but also in practically applying medical terminology (i.e., content), which is, as the researchers pointed out, another language in itself.

The results of the abovementioned qualitative studies assisted in creating a better understanding of the context and classroom environment for using EMI. While the overall perceptions and attitudes towards EMI were positive, many instructors and students were forced to establish strategies to cope with its implementation, raising

---

[8] Adapted from Folkman and Lazarus (1988).
[9] Confrontive coping indicates a degree of risk-taking.
[10] Distancing underlines self-detachment to reduce the significance of a situation.

serious concerns about its use. Since there are many variables that could contribute to the success or failure of implementing EMI (e.g., students' ages, learning styles, classroom discourse, selection method, school type, L1 and L2 background, and so forth), one main recurring variable is L2 proficiency, as both students' perceptions of EMI and comprehension of content depended heavily on their English proficiency level. If students had a sufficient command of English, instructors and students would not need to utilize the coping strategies found by Shamim et al. (2016). None of these studies tested the participants on their language proficiency, and the studies were not able to empirically show a correlation between language proficiency and the outcome of content and/or language learning outcomes. Other studies are needed to "examine empirically the cost and benefits of the use of EMI at HEIs [Higher Education Institutions]; the main goal being how much language is being gained by such programs, as well as how much academic content is being achieved" (Shohamy, 2013, p. 203) or lost. Besides, these studies do not address the type of input instructors use in EMI contexts or to what extent it assists students in comprehending the subject matter. The following section discusses input modification and its influence on learning outcomes.

1.3 Input modifications

As previously mentioned, EMI lacks a clear theoretical foundation or specific objective for language learning. Presumably, EMI can support language learning by providing opportunities for practice and increases language exposure in a meaningful context (i.e., content lessons). Some scholars argue that exposing learners to a large amount of input will gradually improve their language ability; however, it is equally important to assess the *quality* of input used in L2 settings. Researchers who have explored language input have considered certain variables, such as individual differences, learners' awareness, classroom interaction, available feedback, and

pedagogical method (e.g., Krashen, 1981, 1982; Long, 1980, 1985; Schmidt, 1990, 1993, 2001; VanPatten, 1990).

Although researchers rarely define input, it is commonly understood as consisting of "events affecting the visual and auditory perceptual systems" (Carroll, 2001, p. 8). Learners in some situations "attend to some stimulus in the speech environment, process it, and acquire some bit of knowledge about the L2" (p. 9). The same learners in a different situation, however, may not home in on the relevant stimulus and would not grasp anything about the language (Carroll, 2001; Gass & Selinker, 2001). These two situations were distinguished by Corder (1967) as intake and input, respectively. The aim in SLA is to have learners process the stimulus at some level—to mentally represent it—in order to acquire it (Corder, 1967; Krashen, 1981; VanPatten, 1993).

Researchers of first language acquisition have investigated the input modification employed by adults, who normally simplify their speech when speaking to children (e.g., Drach, 1969; Granowsky & Krossner, 1970; Slobin, 1971; Snow, 1972). Based on these findings, SLA researchers hypothesize that teacher-student interactions in language classrooms would also involve some input modification (e.g., linguistic complexity level). Studies on both first and second language acquisition have shown that the degree of linguistic modification is "geared to either the changing age or the increasing skills" (Gaies, 1977, p. 205) of the child or L2 learner. In other words, a more proficient learner will receive less modified input, while lower proficiency learners will receive more modified input.

Modifications of three types of discourse are explored in the literature: literary, natural, and classroom. While literary modification refers to employing a simplified register in print, natural and classroom modifications often refer to a

simplified linguistic register used by a NS conversing with a low-proficiency NNS (Hallett, 2000). It is important to note that, within the print and classroom discourse, there are natural and pedagogical contexts (Beck, McKeown, & McCaslin, 1983). While simplified input often occurs naturally between native and non-native speakers, elaborated input is usually prepared for pedagogic purposes (Kim, 2006, Oh, 2001, Long & Ross, 1993). Simplified speech used in the classroom is the category that relates specifically to the objective of the present study.

In observing L2 teacher-student speech in the classroom, Ferguson (1971) was one of the earliest researchers to investigate "linguistic simplicity" (Gass & Selinker, 2001, p. 260). He stated that native speakers (NSs) tended to adjust their speech when directed to less proficient individuals, exhibiting a style he called "foreigner talk" (Ferguson, 1971, p. 143). He also noted that NSs displayed various means of altering speech to non-native speakers (NNSs), such as speech rate (i.e., a slow rate and longer pauses), word choice (i.e., more frequent words), syntax (i.e., short and simple utterances), and discourse (e.g., provision of feedback).

Generally, NSs adjust their speech to make input comprehensible to L2 learners (Krashen, 1981, 1982). In a series of studies, Chaudron (1983a, 1983b) examined how modifications in spoken discourse affected comprehension and found that repetition of simple nouns helped L2 learners to recall those nouns. He argued that saliency—the noticeability of particular words or phrases in input—is a supportive input feature. These modifications, however, do not consistently appear. Chaudron (1988) and Gass (1997) referred to a number of studies that reported little to no modification by NS participants. Indeed, NSs vary in the type of modification they provide and in their ability to engage with NNSs using modified input, "depending on their communicative style or skills and their prior experience of

communicating with NNSs" (Ellis, 2008, p. 214). Several SLA studies have explored

the effect of linguistic input on language learning (e.g., Kleifgen, 1985; Long, 1983a,

1983b; Wesche & Ready, 1985). According to Chaudron's (1988) thorough review,

many studies during the 1970s and '80s explored the kind of input modification

known as "teacher talk." While these studies provided mixed results because of

confounding variables, they showed overall consistent "intra-subject tendencies to

simplify when telling stories to less proficient listeners" (Chaudron, 1988, p. 68). That

is, when speaking to low-proficiency learners, NSs tend to "downsize" or "water

down" their speech on a variety of levels: phonology (e.g., speech rate and pauses),

lexis (e.g., lexical variety and complexity), syntax (e.g., length of utterance,

subordination, markedness, and grammaticality), and discourse (e.g., framing moves

and self-repetition).

Two major types of modifications have been explored consistently in the

classroom speech of NSs to NNSs: lexical and syntactic. Studies exploring lexical

modification have reported that NSs tend to use basic and more frequent words in

their speech to L2 learners. The utterances carried fewer colloquial instances, fewer

idioms, and more proper nouns (Chaudron, 1982; Henzl, 1973, 1979; Kleifgen, 1985;

Mizon, 1981). The most common measure used to explore lexical variety, as noted by

Chaudron (1988), is the ratio of the number of different words to the number of words

produced: the type-token ratio (TTR[11]). Most studies showed that the TTR of NSs

decreases when they speak to low-proficiency learners (i.e., NSs use less varied

vocabulary).

Syntactic modification studies, on the other hand, have shown conflicting

findings with regard to length of utterance. Whereas some studies have found that

---

[11] A smaller type-token ratio indicates less diverse vocabulary.

NSs use shorter utterances when addressing NNSs, others have found an equal length of utterance regardless of the interlocutor. These conflicting results can be attributed to the different analytical methods, e.g., the unit of analysis employed (T-unit, C-unit, utterance, and sentence; Chaudron, 1988). Other studies measured subordination—the mean number of clauses per T-unit—to assess level of syntactic complexity. Studies that used T-units were somewhat more consistent than others and exhibited similar rates of NSs using shorter utterances with less-proficient NNSs and longer utterances with fellow NSs (Early, 1985; Gaies, 1977; Hakansson, 1986; Henzl, 1979; Ishiguro, 1986; Mannon, 1986; Milk, 1985). Overall, studies indicated that NSs tend to adjust their syntactic complexity "downward" (Chaudron, 1988, p. 78) when addressing less proficient learners (Early, 1985; Gaies, 1977).

While studies of input modification boomed 30 to 40 years ago, fewer recent studies have explored this phenomenon (e.g., Hallett, 2000 [vocabulary acquisition]; Ivanova, 2011 [phonological modification]; Jackson, 2014 [repetition and awareness]). Three key studies that explored input modification are reviewed here because their designs are relevant for the current study (Gaies, 1977; Lynch, 1987; Wesche & Ready, 1988).

An early study by Gaies (1977) investigated the syntax of eight teacher-trainees, of whom three were advanced English NNSs and five were English NSs with no prior teaching experience. All the teachers were taking a practicum course at the time of data collection and collaborated in teaching the classes. Gaies collected two types of data: EFL and baseline. While the EFL data were collected based on three observations of classes taught by each instructor, the baseline data were collected from meetings in which the teachers had to discuss general issues in teaching English. Thus, the baseline data were input from the teachers, who presumably showed equal

proficiency, regarding the practicum course. Based on a T-unit analysis of transcribed sampled data, the results showed that there was a significant "relationship between the syntactic complexity of the subjects' classroom language and the level of proficiency of their students" (Gaies, 1977, p. 209). That is, the teachers' syntax became more complex when geared towards advanced learners and less complex when geared towards low-proficiency learners. Furthermore, low-proficiency learners also received more time to process the input. In most studies exploring input modification, instructor participants are NSs, able to adjust their linguistic input based on students' level of proficiency. This study showed that advanced NNS instructors are also capable of modifying their speech towards low-proficiency L2 learners. It was not clear, however, how proficient the NNS instructors were and whether their modification was based on their language teaching experience.

In a laboratory study, Lynch (1987) investigated the effects on comprehensibility of the modifications made by English NSs when speaking to NNSs. In this study, 24 EFL teachers (NSs) were given three picture-sequence stories of six pictures each and asked to tell the stories to four types of listeners: NSs, and advanced-, intermediate-, and elementary-level NNSs. Each NS sat face-to-face with one listener at a time in the above-mentioned order, and each session was videotaped. Each listener was given a jumbled series of pictures to put in order based on his or her understanding of the story and encouraged to actively interact with the speaker. The study showed three types of modifications: input, interaction, and information choice. Whereas input modification included lexical simplification, such as the use of high-frequency words and avoidance of idiomatic expressions, interaction modification consisted of frequent pauses and comprehension checks, and was primarily used with elementary-level NNSs. Information choice modification involved NSs supplying

vivid details, displaying explicit markers of logical development, and filling in assumed socio-cultural gaps. That is, NSs modified their narrative based on their perceptions of listeners' proficiency and comprehension.

This study showed input modification used when NSs narrate stories to learners with different proficiency levels. However, Lynch's study had seven major drawbacks that the current study has taken into consideration. Besides (a) the absence of an English proficiency measure and (b) the lack of a randomized order of listeners, Lynch (c) did not mention the hypotheses he was testing. In addition, he only descriptively discussed the transcript excerpts (d) with no quantification (e.g., without analyzing the transcripts for lexical diversity, lexical sophistication, and syntactic complexity) or statistical analyses of any kind. In addition, (e) the study's methods and procedure were unclear (e.g., how did the speaker identify the proficiency level of the listener?), and (f) there were no precise measures of the listeners' content and language comprehension. Additionally, (g) it was not ideal to judge listeners' comprehension on the basis of their ability to sequence jumbled pictures correctly, as listeners could have guessed the right order. These drawbacks were avoided in designing the experiment to be reported in this dissertation (see Method section).

Wesche and Ready (1985) explored input modification in a university setting using two languages: English and French. There were two instructors: one English NS and one French NS. Each instructor taught two courses in his or her native language, one oriented towards L2 learners and the other towards NSs. That is, each instructor served as "his or her own control in comparison of discourse to nonnative and native speakers" (p. 89). Each instructor was video-recorded teaching parallel lectures to L1 and L2 sections, capturing approximately the same subject matter in four lectures in psychology (French L1, French L2, English L1, and English L2). The researchers

explored different linguistic aspects in comparing each instructor's lectures for the L1 and L2 groups. Overall, the two instructors showed "dramatically different ranges of variability even within their native speaker-directed speech, as well as across the L1 and L2 situations" (p. 107). Variations were documented in speech rate, nonverbal behavior, and syntactic features. While the English NS instructor showed significant differences between his input to L1 students and L2 students in speech rate and syntactic features, the French NS instructor showed no differences between the two situations. Finally, the researchers claimed that the students successfully learned the subject matter (in both French and English) based on their performance on weekly quizzes and their final exam grades. However, this is not necessarily the case, as the researchers stated that this success could be "partially attributable to their reading" (p. 108). Several factors contributed to the variability seen in the results, including students' prior, or outside, knowledge (e.g., reading), L2 proficiency level, instructor teaching style, and English versus French language-specific norms. Also, the instructors and content were confounding variables that would have been avoided by observing more than one teacher for each language and using more lessons. Furthermore, it was not clear what the proficiency level of L2 learners in each class was, or whether the instructors recognized the learners' level of proficiency.

Finally, as seen in the aforementioned studies, Chaudron (1988) noted different confounding variables that might constrain NSs' input modification, such as NSs' teaching experience, two-way interaction (e.g., instructors getting a sense of the learners' level of interaction), learners' L2 proficiency, content knowledge, and setting (e.g., formal academic situation or informal conversation). While these studies showed evidence of both lexical and syntactic input modifications, research on the effects of these types of modifications on L2 learners' comprehension has been

limited. Indeed, Long and Ross (1993) argued that input modification that removes complex linguistic items leads to less natural forms of the language. Lexical and syntactic modifications could, therefore, deny learners the opportunity to advance to higher proficiency levels (Honeyfield, 1977; Meisel, 1980). While scholars debate the effectiveness of input modification in developing L2 learning, language instructors still modify input based on their perceptions of L2 learners' proficiency level, and instructors in content-and-language-integrated settings, such as EMI and CLIL, most likely modify theirs, as well (Long et al., 2018). One of the questions motivating the current study is whether EMI leads advanced NNS instructors to employ input modification and, if so, what their effects on content learning are.

1.4 EMI and input modification

L2 proficiency level is a recurring key variable in the literature on EMI and input modification. Low proficiency both causes students to struggle in an EMI setting and triggers instructors to modify their input for those students. In contrast, advanced NNSs comprehend more content through L2 input with little to no modification and are more likely to seek more input from resources in and out of the classroom setting (Seliger, 1977). Presumably, EMI aims for effective content learning, although whether students show compelling content learning outcomes through EMI is rarely investigated. A rational hypothesis is that learners' level of language proficiency determines content learning outcomes. DeKeyser (2010), in a study-abroad context, stated that "students must have adequate basic knowledge of the structure of the language if they are to have any meaningful practice experience abroad" (p. 90). Similarly, in an EMI setting, learners with minimal language proficiency will not fully comprehend the content: "It is a given that without understanding the language, no learning can take place. Although understanding alone does not guarantee that learning will take place, it does set the scene for learning to

take place" (Gass & Selinker, 2001, p. 264). In other words, although instructors

might tend to simplify their input in order to improve learner comprehension, doing so

does not necessarily mean that learners take in the subject matter. In fact, instructors'

input modification could lead to content dilution (i.e., loss of essential information

and key lexical items). Few studies have alluded to the idea of content dilution or poor

content delivery, particularly when participants teach, or speak to, learners with low-

proficiency in the medium language (Long & Ross, 1993; Lynch, 1987; Mackay,

1993). With no relation to CLIL or EMI contexts, these studies explored teacher-

student discourse through various methods—e.g., text modification, laboratory study,

and quasi-ethnographic study. Long et al. (2018) study, however, simulated CLIL

contexts and investigated teachers' input for content dilution, along with lexical and

syntactic measures.

     Long et al. (2018) explored the impact of input modification on both content

and language outcomes in a CLIL setting. The study was conducted as a controlled

laboratory experiment simulating a tertiary CLIL setting in Europe, with English as

the medium of instruction. The study had three conditions: CLIL, Arabic baseline,

and English baseline. Each condition had three groups, each consisting of one teacher

and four students, for a total of nine groups, and the CLIL teachers and students were

tested for their English proficiency. The same procedures were followed separately

for every group. Teachers were asked to deliver a 15-minute lesson on fictional

anthropology subject matter (fictional, so as to preclude the possibility of prior

content knowledge) to assigned students. While the Arabic baseline group used

Arabic to provide the lesson, the English and CLIL conditions' lessons were delivered

in English. Two outcome measures were used to test the students' content and

language outcomes: a multiple-choice test and a cloze test. Finally, lessons were both video-recorded and transcribed.

Despite the small sample, the results obtained in this study were key in bridging the research gap between input modification, content dilution, and content-and-language-integrated settings. For each outcome measure, both the English and Arabic baseline conditions outperformed the CLIL condition on the cloze test of vocabulary knowledge, and the English baseline groups outperformed the CLIL groups on the multiple-choice test of subject-matter learning, suggesting greater content comprehension by the English group. With regard to the types of modifications exhibited, instructors in the baseline conditions exhibited greater syntactic complexity than in the CLIL condition, measured by S-nodes per clausal utterance. Likewise, instructors showed less lexical modification in the baseline conditions than in the CLIL condition. As for the content, there was no statistical difference found between the English baseline condition and the CLIL condition in the number of mentioned lexical items; however, the number of times a target word appeared in the instructors' speech was a predictor of the number of correct responses in the cloze test, indicating a direct relationship between the content quality – mentioning the necessary target items – and the students' content learning.

Schierloh and Paulsell (2010) explored input modification in two contexts: a traditional German course and a business German course. Three lessons were video-recorded within a two-week period. The two instructors had taught German for at least 20 years; it was not clear whether they were German NSs. The researchers transcribed the recorded lessons and analyzed the data by marking "episodes" of modified input. They coded salient features for grammar, vocabulary, and stylistic expression, but did not mention the type of measures they used (e.g., T-unit or TTR).

The results indicated that the traditional German instructor used more modification (90% of lessons) than the business German instructor (50%). The business content data involved use of high-frequency words, reduced syntax, and increased use of repetitions. The researchers concluded that business language is not flexible and as such requires the instructor to use certain vocabulary items and phrases in specific contexts. If this was the case for business content, it is likely that medical and science content would, in turn, show similar outcomes. Therefore, replacing specific vocabulary items with simpler, more frequent words could affect the quality of the content.

The results of the aforementioned study are difficult to generalize because the groups (traditional vs. business) were not comparable: the courses had different objectives, students, and instructors. Besides the researchers' lack of a systematic method for analyzing the transcripts, the learners' German proficiency and content learning (i.e., comprehension) were not assessed, and it was not clear how familiar the instructors were with the students' proficiency level. It is possible that the business L2 learners were more advanced than those in the traditional course, leading the business instructor to use fewer modifications. Despite the study's limitations, the results indicated that content-and-language-integrated courses can include some form of input modification. However, the magnitude of such modification in influencing content learning is still undetermined.

Building on the aforementioned studies, the current research was designed to explore three key issues through a controlled laboratory experiment, with implications for EMI contexts: (a) advanced NNSs' and NSs' syntactic and lexical modification, (b) content dilution, and (c) learners' content comprehension. These issues were

investigated while avoiding the confounding factors in the earlier literature mentioned above.

1.5 Linguistic complexity measures

In SLA studies, linguistic complexity in oral and written discourse indicates L2 proficiency. However, there are several definitions and types of linguistic complexity, which have generated different approaches to interpreting results and have caused confusion within the field (Bulté & Housen, 2012; Bulté & Housen, 2014; Norris & Ortega, 2009; Pallotti, 2009, 2015). Currently, there are several measures of L2 complexity in use. Their primary purposes are: "a) to gauge proficiency, (b) to describe performance, and (c) to benchmark development" (Ortega, 2012, p. 128). Additionally, studies have examined input modification using similar measures to those previously listed (TTR, S-units per clausal utterance, etc.). Linguistic complexity can be divided into two main kinds: lexical complexity and syntactic complexity.

1.5.1 Lexical complexity

Lexical complexity has two critical dimensions that have been consistently explored: lexical sophistication and lexical diversity. *Lexical sophistication* refers to the percentage of low-frequency words found in a given text or speech sample (Pallotti, 2015; Laufer & Nation, 1995), with the expectation that the usage of low-frequency lexical items indicates the possession of a larger lexicon (Bulté & Housen 2014). *Lexical diversity*, or variation, refers to "the range of different words used in a text, with a greater range indicating a higher diversity" (McCarthy & Jarvis, 2010, p. 381). Jarvis (2013) added that it also refers to "the proportions of words in a language sample that are not repetitions of words already encountered" (p. 11). SLA research has shown that lexical diversity indices can predict vocabulary knowledge and language proficiency (Bulté & Housen, 2014).

31

Given the multitude of measures for lexical complexity, it is beyond the scope of this dissertation to discuss all of them (for reviews see: Lu, 2012; Jarvis, 2013; Bulté & Housen, 2012, 2014). This study used both lexical diversity and sophistication measures generated by the Lexical Complexity Analyzer (LCA; Lu, 2012). Testing 25 different metrics proposed in SLA, Lu (2012) used the LCA to produce an automated, consistent measurement of lexical complexity's critical dimensions. In a repeated-measures design with test scores, Lu (2012) used the LCA to examine oral test data from Chinese English-learners of varying proficiency levels. He then explored how each of these 25 measures compared to one another as indices of the quality of learners' language production. He found that 13 of 20 lexical diversity measures and two of five lexical sophistication measures showed significant correlations with learners' proficiency, with verb sophistication measures showing particularly robust correlations.

Among the lexical diversity measures, Lu (2012) found that the original TTR had no significant correlations with learners' proficiency in all 12 groups. While the original TTR was not statistically significant, transformed TTRs, such as Corrected TTR (CTTR) and Root TTR (RTTR), were predictors of language proficiency and showed stronger effects when calculated in the LCA. Transformations like CTTR are used to reduce sample size effects, as the original TTR formula is strongly sensitive to sample size, with increased sample size leading to a decrease in the ratio. Proposed by Carroll (1964), CTTR allows the number of word tokens to be adjusted, putting more emphasis on the number of word types in a given text. The ratio is calculated by using the number of word types over the square root of word tokens ($\mathbf{T}/\sqrt{\mathbf{2N}}$).

Verb sophistication also showed a stronger correlation with learners' proficiency than non-transformed TTR. This measure was originally introduced by

Harley and King (1989), who defined it as the ratio of the number of sophisticated, advanced verb types to the total number of verbs employed in a given text or speech (e.g., number of sophisticated verb types/total number of verbs[12] – **T**sverb/**N**verb). Harley and King (1989) identified verbs as "sophisticated" if they were not on the list of the 200 most frequent French verbs. In the LCA,[13] verbs were considered "advanced" if they were not part of the 2,000 most frequent words generated by the British National Corpus (BNC; Leech, Rayson, & Wilson, 2001). Wolfe-Quintero, Inagaki, and Kim (1998) adapted Carroll's (1964) CTTR formula to reduce the sample size effect in extracting lexical sophistication. Thus, Wolfe-Quintero et al. (1998) proposed using Corrected Verb Sophistication-1 (CVS1) to minimize sample size effect. CVS1 is a transformation of the original VS1 (Harley & King, 1989) and is calculated using the number of advanced verb types over the square root of verb tokens (**T**sverb/$\sqrt{2Nvreb}$). Both CTTR and CVS1 were used in this study (for details, see Materials).

1.5.2 Syntactic complexity

The second type of linguistic complexity, syntactic complexity, refers to "the range of forms that surface in language production and the degree of sophistication of such forms" (Ortega, 2003, p. 492). That is, syntactic complexity indicates the level and range of syntactic sophistication produced by the learners, in both oral and written forms (Crossley & McNamara, 2014; Lu, 2012). Measures of syntactic complexity primarily aim to "quantify one or more of the following: range of syntactic structures, length of unit, degree of structural complexity ('sophistication') of certain syntactic

---

[12] Also, refers to as verb tokens
[13] LCA treats different inflections of the same lemma as one type (e.g., "eat," "eats," "eating," "ate," and "eaten").

structures and amount and type of coordination, subordination and embedding" (Bulté & Housen, 2012, p. 35).

Many syntactic complexity measures have been used primarily to evaluate learners' writing. The most common traditional measure is the T-unit, which Hunt (1970) defines as "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it" (p. 4). The T-unit determines syntactic complexity in written forms, in that it measures the effect of sentence combining, and it has been used heavily in the literature (Gaies, 1980). However, research using T-unit measures in assessing L2 writing has shown inconsistent results (Crossley & McNamara, 2014). Measures such as length of syntactic structures, types of coordination between clauses, and the frequency of clauses have also been employed to examine syntactic complexity in L2 writing (Crossley & McNamara, 2014), most of which are designed to evaluate learners' overall writing development.

While research on syntactic complexity in teachers' speech has been limited, researchers have commonly applied T-unit indices to assess learners' oral language development (Crossley & McNamara, 2014). However, this study uses another way of measuring oral data: the ratio of Sentence-nodes, or S-nodes, to clausal utterances (see Materials for detailed definition). The number of S-nodes per clausal utterance is a better indicator of syntactic complexity in speech production than the T-unit, which measures written language, and avoids "the extensive reduction of raw data necessitated by a T-unit analysis" (Crookes, 1990, p. 189). Given the spoken discourse, S-nodes per clausal utterance was also clearly operationalized and proved to be a reliable measure in comparison to other measures. This measure was also employed by Long et al. (2018) to examine the syntactic complexity of teachers' speech.

Chapter 2: The study

2.1 Purpose and objectives

This study was motivated by five gaps in the SLA literature on EMI and input modification in L2 classrooms.

1. Numerous studies that have examined EMI globally, and the few conducted in the context of Saudi Arabia, have depended on qualitative methods (questionnaires, interviews, and case studies) to explore students' and instructors' attitudes towards EMI. These studies neglected to discuss the unsystematic implementation of EMI, the debate on whether EMI is considered a pedagogical method, language use in EMI, and EMI's effectiveness in language and content learning.

2. Many of the studies that have examined input modification in L2 classrooms and in one-on-one settings (e.g., in conversation or narrating stories) lack important methodological features, such as quantification of results, measures of learners' language proficiency, measures of comprehension, and so forth. While these studies generally document the existence of input modification, the quality and quantity of modifications (e.g., lexical and syntactic modifications) are still unclear, and the lack of empirical measures has led to mixed results.

3. In the input modification literature, few studies have explored the effects of modifications on learners' content comprehension.

4. Little research has investigated whether advanced NNSs modify input when teaching content or narrating stories to less proficient listeners or compared their input modification to those of NSs.

5. Very few studies have investigated whether input modification leads to content dilution (e.g., omission of information bits). Knowledge of the impact of such loss of content on learners' comprehension is limited.

Considering these five gaps, this study's main goals are (a) to use a laboratory setting with many of the characteristics of EMI to investigate input modification by (b) comparing the speech of advanced NNSs towards low-proficiency learners of English to that of NSs (c) using quantitative methods (i.e., measures of lexical complexity and syntactic complexity). In addition, the study explores (d) the effect of these modifications on learners' content comprehension and (e) the possible content dilution triggered by the input modification (f) using empirical measures.

To address these objectives, the study examines types of input modification and content dilution exhibited by English NSs and NNSs when narrating a story to three different types of listeners: native English-speakers (ESs),[14] high-proficiency learners, and low-proficiency learners. The experimental design was adapted from Lynch's (1987) study, with five methodological improvements: (a) an English proficiency measure, (b) a measure of content comprehension, (c) a randomized order of listeners per speaker, (d) a randomized order of stories per listener, and (e) quantitative analyses in examining transcripts. The research questions and hypotheses are outlined in the following section.

2.2 Research Questions and Hypotheses

- **RQ1:** Do NSs of English differ from advanced NNSs in their input modification when narrating the same story in the three listener conditions?

---

[14] In order to distinguish listener and speaker condition of NS, ES is used here to refer to the listener/student and NS is used to refer to the speaker/teacher.

- o **H1a**: NSs will use more lexically diverse speech (i.e., fewer word repetitions) than NNSs in the three listener conditions. This difference will be indicated by a higher CTTR value for NSs.

- o **H1b**: NSs will use more lexically sophisticated speech (i.e., more low-frequency words) than NNSs in the three listener conditions. This difference will be indicated by a higher CVS1 value for NSs.

- o **H1c**: NSs' speech will exhibit greater syntactic complexity than NNSs' speech in the three listener conditions. This difference will be indicated by a higher number of S-nodes per clausal utterance for NSs.

- o **H1d**: NSs will include more content details in their speech than NNSs in the three listener conditions. This difference will be indicated by higher counts of mentioned information bits (IBs) by NSs.

- **RQ2:** How do speakers *with* language teaching experience (WLT) compare to speakers with *no* language teaching experience (NLT) in terms of input modification when narrating the same story to the three listener conditions?[15]

  - o **H2a**: Speakers WLT will use more lexically diverse speech in the three listener conditions than speakers with NLT. This difference will be indicated by a higher CTTR value for speakers WLT.

  - o **H2b**: Speakers WLT will use more lexically sophisticated speech in the three listener conditions than speakers with NLT. This difference will be indicated by a lower CVS1 value for speakers WLT.

  - o **H2c**: Speakers WLT will exhibit more complex syntax in their speech in the three listener conditions than speakers with NLT. This difference

---

[15] There were no studies, to this date, that investigated input modifications for speakers WLT in comparison to speakers NLT; thus, the hypotheses regarding the influence of LT experience were based on the researcher's reasoning. Provided their language teaching experience, speakers WLT will generally have higher values, and wider range, in complexity compared to speakers NLT.

will be indicated by a lower number of S-nodes per clausal utterance for speakers WLT.

- o **H2d**: Speakers WLT will include more content details in their speech in the three listener conditions than speakers with NLT. This difference will be indicated by higher counts of mentioned IBs for speakers WLT.

- **RQ3:** How do NSs modify their input according to listener proficiency when narrating the same story?

  - o **H3a**: NSs will use less lexically diverse speech with low-proficiency listeners. This difference will be indicated by a lower CTTR value for low-proficiency listeners.

  - o **H3b**: NSs will use less lexically sophisticated speech with low-proficiency learners. This difference will be indicated by a lower CVS1 value for low-proficiency listeners.

  - o **H3c**: NSs will exhibit less complex syntax in their speech with low-proficiency listeners. This difference will be indicated by a lower number of S-nodes per clausal utterance for low-proficiency listeners.

  - o **H3d**: NSs will omit more content details in their speech with low-proficiency listeners. This difference will be indicated by lower counts of IBs mentioned for low-proficiency listeners.

- **RQ4:** How do advanced NNSs modify their input according to listener proficiency when narrating the same story?

  - o **H4a**: NNSs will use less lexically diverse speech with low-proficiency listeners. This difference will be indicated by a lower CTTR value for low-proficiency listeners.

- o **H4b**: NNSs will use less lexically sophisticated speech with low-proficiency learners. This difference will be indicated by a lower CVS1 value for low-proficiency listeners.

- o **H4c**: NNSs will exhibit less complex syntax in their speech with low-proficiency listeners. This difference will be indicated by a lower number of S-nodes per clausal utterance for low-proficiency listeners.

- o **H4d**: NNSs will omit more content details in their speech with low-proficiency listeners. This difference will be indicated by lower counts of IBs mentioned for low-proficiency listeners.

- **RQ5**: What effect do input modification and content dilution have on the listener comprehension?

  - o **H5a**: High-proficiency listeners will exhibit greater comprehension of the stories than low-proficiency listeners. This difference will be indicated by higher content comprehension assessment scores.

  - o **H5b**: ESs listeners will exhibit greater comprehension of the content than both low- and high-proficiency listeners. This difference will be indicated by higher content comprehension assessment scores.

Table 2 offers a summary of what each measure means in relation to different type of hypotheses.

Measure	Hypothesis	Ratio/counts	Meaning
CTTR	More lexical diversity	High ratio	Fewer repetitions
CTTR	Less lexical diversity	Low ratio	More repetitions
CVS1	Higher lexical sophistication	High ratio	More low-frequency words
CVS1	Less lexical sophistication	Low ratio	More high-frequency (common) words
SC	Higher syntactic complexity	High ratio	More S-nodes per clausal utterance
SC	Less syntactic complexity	Low ratio	Fewer S-nodes per clausal utterance
IB	Omit more content details	Low counts	Less mentioned information
IB	Omit fewer content details	High counts	More mentioned information

Table 2: Outcome measure interpretations

Chapter 3: Methodology

3.1 Participants

This study used a criterion-group design with two types of participants: speakers and listeners. Two types of speakers were recruited: NSs and advanced NNSs of English. Each group consisted of four male and six female participants (ages 26-40). Participants were selected based on academic background and/or teaching experience. The advanced non-native speakers were native Arabic-speaking graduate students (in at least the third or fourth year of their programs) at an English-speaking university, with a score of at least 6.5 on the IELTS. The English NSs had a similar academic level to the NNSs. The speaker participants were further divided into two subgroups: those with language teaching experience (WLT) and those without (NLT). Language teaching was a moderator variable because many, if not all, studies exploring input modification include language teachers as participants; therefore, investigating the influence of language teaching experience on input modification would shed more light on the topic. The ten speakers with NLT had absolutely *no* language teaching experience, and their degrees were not related to language studies; the ten speakers WLT were either currently teaching or had at least a year of experience as language teachers. The researcher followed up with each participant to schedule a one-on-one study session. The 20 speakers were assigned to one of four subgroups of five participants each: NSs WLT, NSs with NLT, NNSs WLT, and NNSs with NLT.

For the listeners, three groups of participants were recruited: ESs (control), high-proficiency NNSs, and low-proficiency NNSs. There were 60 listeners in total (20 ESs, 20 high-proficiency learners, and 20 low-proficiency learners). All listener participants were recruited from a large US public university through emails, flyers,

and word of mouth. The non-native listeners were recruited from an intensive English communication program and from among local native Arabic-speakers. Thirty-two of the participants were from Saudi Arabia, and eight from other Arabic-speaking countries (three from Kuwait, two from Oman, and one each from the UAE, Iraq, and Jordan). Both non-native groups were given a dictation test as a measure of their listening proficiency before their assigned sessions were scheduled.

The high-proficiency listeners were placed in either level three or four (intermediate to high) of the intensive English program. They had not achieved the minimum IELTS score, 6.5, required by many US universities for international student admission; therefore, their level in a real-world context would be considered intermediate. In this study, they were labeled "high-proficiency" simply to distinguish them from low-proficiency participants. The low-proficiency learners were either high school graduates (mainly from Saudi Arabia[16]) who had come to the US for reasons related to family, work, or education,[17] or students in level one or two (beginners) in the intensive English program.[18] The high-proficiency group consisted of twelve male and eight female participants (ages 18-32), and the low-proficiency group of eight male and twelve female participants (ages 18-35). The non-native listener groups consisted primarily of undergraduate and graduate students. The ES control group also consisted of undergraduates and graduates – five males and fifteen females (ages 18-30). At the end of each session, the speaker participants were compensated with $20 for their participation (total of $60 for three sessions), and the listener participants with $10 for their participation.

---

[16] For a reminder of the Saudi educational context, refer back to Chapter 1.
[17] Some are BA holders who sought their degrees (unrelated to language studies) from universities in Saudi Arabia or in another Arabic-speaking country.
[18] Some of the non-native listeners were in the intensive English program to assist them in pursuing a master's degree in an English-speaking university.

3.1.1 The proficiency measure

Participants' proficiency was measured by administering an English dictation test. In dictation, a script is read aloud to learners, who attempt to accurately transcribe what they hear. Dictation has been used in language classrooms for decades (Oller, 1971, 1979; Sawyer & Silver, 1972; Morris, 1983; Fisher, 2001; Nation, 2009), and researchers have suggested that it be used as a language proficiency measure. Oller (1979) indicated that dictation works well because "the whole family of auditory tasks that it comprises faithfully reflect crucial aspects of the very activities that one must normally perform in processing discourse auditorily" (p. 266). During a dictation test, students must simultaneously receive and process auditory information in order to produce meaningful writing (Oller, 1971). Because this study's central task required listening, a dictation test seemed the most appropriate measure of the listeners' English proficiency.

The dictation script was adapted from Heaton (1966) and based on a picture-sequence story (see Appendix B). This script was chosen because it resembled the type of materials used in the study. The script was then broken into sentences, or, if the sentence was too long, meaningful sequences. A female native speaker of English, who was also an English instructor, recorded the script twice: first, read aloud in a natural voice (about 25% slower than would be used with a NS listener) and second, read aloud at the same pace, with pauses between each meaningful sequence. The pauses between the sentences/sequences were determined based on the length of each sequence and the time it took two NNSs to write down each one (ranging from 12 to 18 seconds). The test was piloted with two NSs to assess the recording and to rule out any possible errors. Then the test was administered to six NNSs of varying proficiency levels to assess the appropriateness of the script and the length of pauses. The scores were 22, 29, 30, 60, 68, and 82, out of a possible 100. The range of scores

was desirable, as it accurately reflected the chosen participants' proficiency levels, as defined by their placement levels in the English program.

To administer the test, the researcher met with the participants (individually or in a group of two or three) in a quiet room before scheduling the study session. They were given paper to write on with the instructions at the top in both English and Arabic (see Appendix A). They were instructed to listen to the recordings carefully and write down as much of what they heard as possible. The script was played twice: the first time without pauses between sentences, and the second with pauses. After the participants finished, their papers were collected and scored. The script contained 100 words, including articles and prepositions, and scores were based on the number of words written down correctly out of a possible 100. Words that were slightly misspelled but still recognizable were counted, but extremely misspelled words – e.g., based only on phonology ("efinng" instead of "evening") or words with different meanings ("cut up" instead of "caught up") – were not counted in the final scores.

Participants who scored 45 words or below were considered to have relatively low proficiency; those who scored 48 words or above were considered to have relatively high proficiency. These score cutoffs were established based on the range of all dictation scores. The low-proficiency participants had a score range of 22 to 45 (*M*: 32 and *SD*: 7.11), the high-proficiency participants scores ranged from 48 to 90 (*M*: 68 and *SD*: 11.7). An independent samples t-test was used to compare the proficiency scores of participants in the two listener conditions, the results yielded a statistically significant difference between the high- and low-proficiency means (t = 21.25, p < 0.001). Both groups of listeners were randomly paired with native and non-native speakers. A one-way ANOVA indicated that there was no difference between

speaker type in terms of listeners' proficiency ($F(1,178) = 0.02$, $p = 0.88$), suggesting that the groups paired with each speaker type were comparable.

3.2 Design

The purpose of the study was to investigate input modifications used by English NSs and advanced NNSs when narrating stories to low- or high-proficiency NNS listeners, as compared to NS controls. Each speaker was randomly assigned to meet with three listener participants in separate one-on-one sessions on the same day. Every speaker met with one participant from each listener group (ES, High, and Low). Each listener met with only one speaker and heard three different stories narrated by that speaker. As shown in Figure 2, every session included three main stories that were presented in randomized order for every session. The speakers' narrations were audio-recorded to examine the types of input modifications employed with high- and low-proficiency listeners, as opposed to ES controls, using measures of lexical diversity (CTTR), lexical sophistication (CVS1), syntactic complexity—i.e., S-nodes per clausal utterance (SC), and content dilution—i.e., omission of information bits (IB). At the end of each narrated story, the listeners were asked to take a content comprehension assessment, referred to here as listener comprehension (LC), and write down, in their native language, what they had understood from the story, providing as many details as possible.
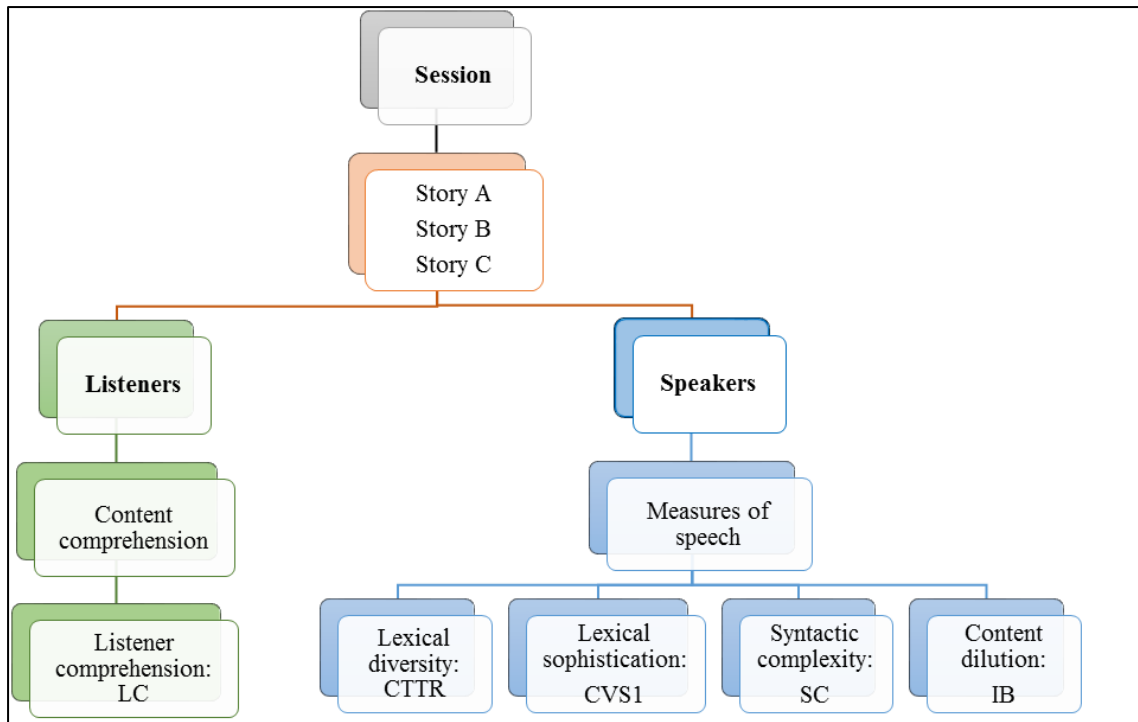
Figure 2: Study Design

3.3 Materials
3.3.1 Stories

Three picture-sequence stories were adapted from Heaton (1966; 1975): *The Elephant Weight* (labeled "A"), *The Blind Man* ("B"), and *The Indian Man* ("C"). The stories were piloted in order to make sure that they (1) elicited enough information from the speakers to allow for variation in what information was included or excluded; (2) were not familiar to the participants, to control for prior knowledge and avoid guessing; and (3) were culturally understandable to listeners (i.e., excluding specifically Western elements that might limit listeners' comprehension)[19]. In addition, two picture-sequence stories were used as a warm-up to prepare both speakers and listeners for the main task. These warm-up stories were also adapted from Heaton (1966; 1975), but no scripts were provided to the speakers before or during the study session (see sample in Appendix C). The listeners were assessed only

---

[19] These three main stories were selected to be mainly understandable, culturally neutral, and domain-general – i.e., not favoring any particular area of study.

on the three main stories. Each main story had an outcome measure that assessed listeners' understanding of the story's content.

The main stories consisted of written scripts given to the speakers before starting their sessions (Appendix D). The stories were reviewed with speakers to avoid the possibility of their misinterpreting some of the frames in the pictures (e.g., the blind man thanking the woman [correct] as opposed to the blind man thanking the boy [incorrect]). As the pilot study had shown, misinterpretation of the pictures would influence listener comprehension and result in uncontrolled variance in the outcome measure, because the comprehension assessment was scored based on the counts of propositions, which were in turn based on the standard scripts. Additionally, in the pilot study, the speakers had omitted many of the stories' details (e.g., "they drove to a forest and passed by a zoo" or "the woman slams the door"), and reading the script ensured that they mentioned more details, to provide a greater possible range among listener conditions.

3.3.2 Linguistic measures
3.3.2.1 Lexical complexity

After the audio recording was transcribed and verified, the transcripts were segmented and separated by story in order to calculate the lexical complexity measures of each story through Lexical Complexity Analyzer (LCA).[20] Each story's text file was appropriately categorized with a session ID, listener type and ID, and a story label. Once all files were completed, they were uploaded and processed, three sessions at a time, in LCA. LCA then generated Excel files that consisted of the measures of lexical diversity (CTTR) and sophistication (CVS1) for each story.

---

[20] The program use is straightforward and has a web-based version developed by Ai and Lu (2010): http://aihaiyang.com/software/lca/

3.3.2.2. Syntactic complexity

The transcripts were also coded for S-nodes per clausal utterance as a measure of syntactic complexity. Following Crookes (1990), S-nodes were defined as tensed or untensed verbs in clauses containing one or more verbs. The inclusion of both verb forms yields a precise measure of syntactic complexity. The use of non-finite verbs with finite verbs is more complex than the use of finite verbs only in a single clause; the clausal utterance "the boy [wants] to [buy] a toy" is more complex (2 S-nodes) than "the boy [wants] a toy" (1 S-node). In coding the transcripts, both tensed and untensed verbs were counted. However, repeated and self-corrected verbs were counted once, and verbs that referred to the speaker (e.g., "I don't know") were not counted.

An utterance is defined as "a stream of speech with at least one of the following characteristics: (1) under one intonation contour, (2) bounded by pauses, and (3) constituting a single semantic unit" (Crookes & Rulon, 1985, p. 9, cited in Crookes, 1990, p. 187). Based on this definition, this study defined a *clausal* utterance as a stream of speech that (1) consists of one semantic unit[21] and (2) has at least one clause[22], (3) under one intonation contour,[23] and (4) bounded by pauses.[24] Unlike a simple utterance, a clausal utterance excludes fragments such as "yes" and "okay." Both grammatical and ungrammatical clausal utterances were included in the coding process – e.g., incorrect subject-verb agreement, missing objects, and incorrect use of articles or prepositions. Also, confirmation checks and utterances irrelevant to the content of the story were excluded from the segmentation.

---

[21] Refers to the same instance/idea.
[22] Fragments are not included.
[23] A falling intonation signals the end of an utterance that often indicates a shift in topic/content.
[24] Pauses must be used in conjunction with either intonation contour or semantic unit in identifying units.

The researcher coded S-nodes and clausal utterances by segmenting each story into several units (utterances), and then counting the verbs (S-nodes) the utterances contained. Units were coded using slashes, and verbs were identified by using double colon (e.g., /the men :: were baffled by a problem / /they :: did not know how :: to weigh a big elephant on a small scale/ – 2 clausal utterances and 3 S-nodes). A second person was trained on this measure and instructed to code a number of transcripts, in order to establish inter-rater reliability. The second rater and the researcher independently coded ten percent of the transcripts.

Inter-rater reliability was assessed using the inter-class correlation coefficient (ICC). ICC was originally proposed as a valid measure of inter-rater reliability by Shrout and Fleiss (1979). ICC is a distinctive measure because it is more flexible than other reliability indices (e.g., numbers of rater and scoring method), and the rating does not have to be a binary integer (0 or 1) in order to run the assessment (Landers, 2015). Using the ICC coefficient with the absolute agreement definition, the inter-rater reliability coefficients for S-nodes (0.986) and clausal utterances (0.930) were high. In gauging the syntactic complexity ratio for all sessions, the mean S-nodes per clausal utterance was calculated for each story by dividing total S-nodes by total clausal utterances. Thus, a transcript with long clausal utterances that contained many S-nodes was considered more syntactically complex than one with shorter clausal utterances and few S-nodes.

3.3.3 Content measures
3.3.3.1 Content dilution
Content was operationalized by the number of propositions mentioned in the story transcripts. Each story contained 26 propositions (78 in total), which included all the main elements (see Appendix F). In coding the transcripts, all propositions – information bits (IBs) – mentioned by the speaker were compared to each story's pre-

determined list. For consistency, additional details mentioned by speakers in response to listeners' questions were not counted. In addition, no points were awarded in cases of absent or wrong subject references (an example of a pre-determined proposition – :: Based on the weight of the stones, :: *the boy* :: calculated the weight of the elephant. (3pts.)). Hence, references to the wrong subject in narrating a story were not counted (e.g., "*they* :: add up the total" [IB: 1] vs. ":: *the boy* :: calculated the total" [IB: 2].) The researcher counted the number of IBs in the transcripts by coding each story for the number of propositions mentioned. Each story received a total IB count at the end of each coding. As with syntactic complexity, inter-rater reliability, through ICC coefficient, was established by first training another person to code the transcripts for the counts of IBs. Ten percent of the transcripts were independently coded by the researcher and the second rater, and agreement was an acceptable 0.936.

3.3.3.2 Content comprehension

Finally, the listener participants were asked to restate the story in writing using their native language (Arabic or English; see Appendix E). Unlike in the study by Lynch (1987), the listeners did not receive a mixed-order sequence of pictures, as this could have led them to guess the right order of the story. Therefore, the listeners had to depend only on the input they received from the speaker. Listeners' comprehension of the stories was scored based on the pre-determined list of propositions (the same one used for the counts of IB; Appendix F). Each story was scored by counting the number of propositions listeners mentioned in their retelling of the story. Ten percent of the listeners' accounts were independently scored by two raters, and an acceptable level, using inter-class correlation, was attained (0.958).

3.4 Procedures

During recruitment, the researcher administered the dictation test to the non-native listeners, as a measure of their listening proficiency. After scoring, the

participants were contacted to schedule a study session based on their level and availability. Then, the listeners were randomly assigned to the speakers – three listeners per speaker – and their one-on-one sessions were scheduled. A day before the scheduled sessions, the speaker and the three listeners received individual email reminders, along with a copy of the consent form and the instruction sheet in their native language (see Appendix G and H for speakers' and listeners' instructions). The speaker was asked to arrive 30 minutes before the first session, and each listener was given a specific session time, in an effort to reduce waiting-time. Depending on proficiency level and writing time, each session took 20 to 30 minutes, for a total of approximately two hours for the speaker. All study sessions were conducted on campus in a laboratory equipped with noise isolation and echo elimination booths. Each booth had at least one table and three chairs for the speaker, the listener, and the researcher.

On the day of the scheduled session, the researcher met first with the speaker in the lab and provided the materials (story scripts, picture sequences, instructions, and a hard copy of the consent form). The researcher gave the speaker scripts of the main stories and reviewed them with him or her to prevent the possibility of misinterpretation. In addition, the warm-up stories were reviewed without scripts, and the researcher emphasized the purpose of using these stories before starting the main stories to get a sense of the listeners' understanding. After describing the study procedure to the speaker and answering his or her questions, the researcher left the speaker to review the stories and instructions until the first listener arrived.

During the scheduled session, the speaker sat face-to-face with one listener at a time in a quiet booth. Each listener was provided with the instruction sheet, content comprehension test, and a hard copy of the consent form, in his or her native

language. The speaker was only permitted to use the picture sequences, along with the instructions. The written scripts were removed as soon as the session started, and the speaker was not permitted to read or review the scripts between sessions. The picture-sequence sheets were used for both warm-up stories and the main stories to assist the speaker in remembering each story's details and narrate the story accurately. The sheets were placed in front of an open laptop screen, such that the speaker could see them while narrating each story, but the listener could not. To avoid unwanted variation in telling the story (such as skipping a proposition or frame), the speaker was instructed to talk about every frame in the picture-guided stories.

Session A Listener 1	Introduction & warm-up	Story 1 LC	Story 2 LC	Story 3 LC
Session B Listener 2	Introduction & warm-up	Story 1 LC	Story 2 LC	Story 3 LC
Session C Listener 3	Introduction & warm-up	Stroy 1 LC	Story 2 LC	Story 3 LC
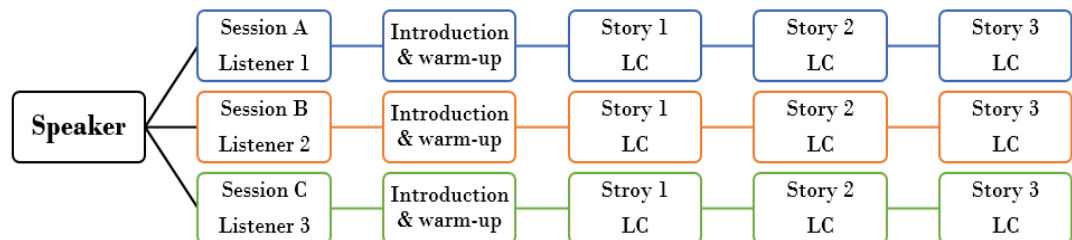
Figure 3: Study procedures

As Figure 3 illustrated, each session comprised three stages: introduction, warm-up stories, and main stories. As part of the introduction stage, the speaker led a three-to-five-minute icebreaker activity, so as to learn the listener's language level. Both participants introduced themselves (i.e., name, hometown, degree/major, and hobbies). Then, the speaker was directed to ask specific icebreaker questions to further engage the listener in a conversation (see Appendix G for instructions for speakers). Then the speaker moved on to the warm-up stories. The speaker narrated a short, three-frame picture-sequence story to the listener. Listeners were encouraged to give feedback and ask questions during the narration; the speaker was instructed to pay attention to the listener's feedback and respond accordingly. Once the speaker finished the first warm-up story, the listener was instructed to retell the story to the

speaker in English. If the listener understood the story, the speaker was instructed to move on to the second warm-up story. If, however, the listener did not understand the story, the speaker needed to explain and correct the parts the listener had got wrong or omitted. In some cases, the listener asked the speaker to repeat the story; this was permitted for the warm-up stage. After the listener successfully retold the story, the speaker moved on to the second warm-up story, which consisted of six frames and included more details, similar to the main stories. The speaker and the listener were instructed to follow the same procedures as with the first warm-up story: the speaker narrated the story, the listener retold the story, and then the speaker provided feedback. The purpose of the introduction and the warm-up stage was to provide the speaker with enough output from the listener to determine his or her proficiency level and whether or not further modifications were needed.

During the main stories stage, the speaker narrated each story, with pictures visible to the speaker, but not the listener. The listener was encouraged to ask questions and request clarification. Once the speaker had finished the first story, the listener was asked to retell it in his or her native language in writing, using the paper provided. The listeners were instructed to include as much detail from each story as possible. After he or she had finished restating the first story, the speaker narrated the second story and then the third, following the same procedure. The researcher was present in each session to monitor and provide necessary guidance. Each session was audio-recorded using a Sony IC recorder (Model: ICD-PX312) and was subsequently transcribed. At the end of each session, the listener was accompanied out of the room and compensated. The second and third listeners scheduled with the same speaker followed the same procedure. Finally, each speaker was compensated at the end of the three back-to-back sessions. Sometimes speakers were asked to come back the next

day, due to the absence of one of the listeners (only occurred for two speakers). In these cases, the speakers had two sessions the first day and one session the next, following the same procedures.

The three stories were used for each listener to preclude the confounding variable of having a single content. They were not controlled for difficulty. The content of each story is evidently different, and the difficulty could vary based on listener vocabulary and speaker narration. Furthermore, in an attempt to obviate a practice effect in the speakers, the stories were presented in a different order for each session, and the listener conditions were counterbalanced. The six possible orders are listed in Table 3. Each order occurred no more than four times, and no two conditions were placed in the same order more than four times. These orders were predetermined and employed randomly when scheduling the order for each study session.

Instructors/ Sessions	First	Second	Third
Order 1	High	ES	Low
Order 2	ES	Low	High
Order 3	Low	High	ES
Order 4	High	Low	ES
Order 5	ES	High	Low
Order 6	Low	ES	High

Table 3: Listener conditions counterbalanced list

Chapter 4: Results

This chapter starts with a discussion of data preparation and story effect and then moves into a presentation of the results for the outcome measures. Linear mixed-effect models (LME) were used to analyze each outcome measure to account for any statistical difference or interaction. For each outcome measure, additional analyses were carried out as needed using analysis of variance (ANOVA) to examine the differences between levels in Speaker Type (NS and NNS) and in Language Teaching (WLT and NLT) conditions. Finally, further LME analyses were conducted within each Speaker Type to investigate differences, within each group, between Listener Types.

4.1 Transcripts and data preparation

Two types of data were collected: transcripts of speaker narration and the listener comprehension assessment. For the transcripts, two process variables were considered: linguistic and content modifications. Four out of five predictors – lexical diversity (CTTR), lexical sophistication (CVS1), syntactic complexity (S-nodes per clausal utterance [SC]), and content dilution (information bits [IB]) – were based on the speakers' transcripts. Therefore, all audio-recordings were carefully transcribed and verified independently by two people. First, the audio-recordings were categorized under labels indicating speaker condition, language teaching experience, listener condition, and ID. Each audio file was reviewed to ensure sound quality and to cut out the introduction and warm-up stage. All audio editing was done using *Audacity*, cross-platform software for recording and editing sounds (version 2.1.2). Then the main stages of the audio files were transcribed using *Transcribe*,[25] an open web application designed to make transcribing audio more practical by incorporating

---

[25] http://otranscribe.com

the audio file and the written transcript in one page. Each audio file included the three main stories and ranged from three to fourteen minutes, for a total of 331 minutes (5.5 hours). All transcripts followed specific transcription guidelines and conventions to assist in representing, as closely as possible, each speaker's syntax (regardless of grammatical errors), words (including word repetitions and self-corrections), and rate of delivery (indicated by pauses; Appendix I). Each transcript was formatted in one table with the speaker in the left column and the listener on the right. Each line was numbered according to story (see Appendix J for a sample transcript).

Each speaker was linked to three sessions: one for each listener type (ES, high-proficiency, and low-proficiency). Each listener session consisted of three stories, for a total of three transcripts per listener and nine transcripts per speaker (3 listeners x 3 stories). In total, there were 180 transcripts (3 stories by 60 listeners). In addition to the transcripts, 180 listeners' content comprehension (LC) assessments for each story were analyzed (3 assessments for 60 listeners). Each outcome measure was assessed through the transcripts and assessments (as explained in the Materials section), which were aligned with each listener condition (Listener Type), speaker condition (Speaker Type), language teaching condition (LT), story, and session identification.

The dataset was analyzed using R (R Core Team, 2013). Each of the five outcome measures – CTTR, CVS1, SC, IB, and LC – was analyzed separately. In cases of skewed distribution, different transformation methods were used, depending on the level of skewness. Assessing the violation of normality assumption depends mostly on two dimensions: the sample size and the shape of the distribution. That is, a distribution considered unacceptable for a small sample size might be unimportant for a larger sample size (Hair, Black, Babin, & Anderson, 2010). The dataset had a total

56

of 180 observations with various groups and conditions. Given the sample size, both the Shapiro-Wilks test of normality and a probability plot were used to assess normality for each outcome measure's distribution.

4.2 Story effect

The three stories were not controlled for difficulty and were initially treated as the same. However, in-depth analysis revealed distinct results for each story that influenced overall results. A one-way ANOVA was conducted to compare the effect of Story on each outcome measure. The results revealed a significant effect of Story on CTTR ($F(2,177) = 22.04$, $p < 0.001$), SC ($F(2,177) = 6.631$, $p < 0.001$), IB ($F(2,177) = 29.5$, $p < 0.001$), and LC ($F(2,117) = 5.767$, $p < 0.001$). However, there was no significant effect of Story on CVS1 ($F(2,177) = 1.019$, $p = 0.363$; see Mean and SD in Table 4). The results indicated that the stories were different in terms of lexical diversity, syntactic complexity, content dilution, and listener comprehension. Story A showed the overall lowest outcome, followed by Story C and then Story B, suggesting that Story A may be the most difficult (Figure 4). All stories had the same number of propositions, but Story A had the most technical details (i.e., importance of location, rational of weighing the elephant, and the boy's suggestion) along with an unconventional concept (i.e., a creative way of weighing the elephant on a barge), indicating that Story A had more complex details than Stories B and C.

In an alternative interpretation, however, speakers may have perceived Story A as the least difficult where they had less cognitive pressure to bring all possible means to bear on their speech (R. DeKeyser, personal communication, April 10, 2018; see *resource-directing* compared to *resource-dispersing* complexity, e.g., Robinson, 2003, 2005; Robinson & Gilabert, 2007). Nonetheless, Story was a major moderating variable for this dataset and was accounted for in further analyses.

Story	CTTR*	CVS1	SC*	IB*	LC*
A (n = 60)	3.74	0.285	2.12	16.9	8.45
	(0.32)	(0.13)	(0.37)	(2.97)	(5.22)
B (n = 60)	4.01	0.287	2.35	20.8	12.46
	(0.34)	(0.17)	(0.44)	(2.21)	(5.72)
C (n = 60)	4.13	0.321	2.09	18.5	10.80
	(0.31)	(0.16)	(0.44)	(3.14)	(4.56)

* showed a significant effect of Story

Table 4: Descriptive statistics of each measure per Story
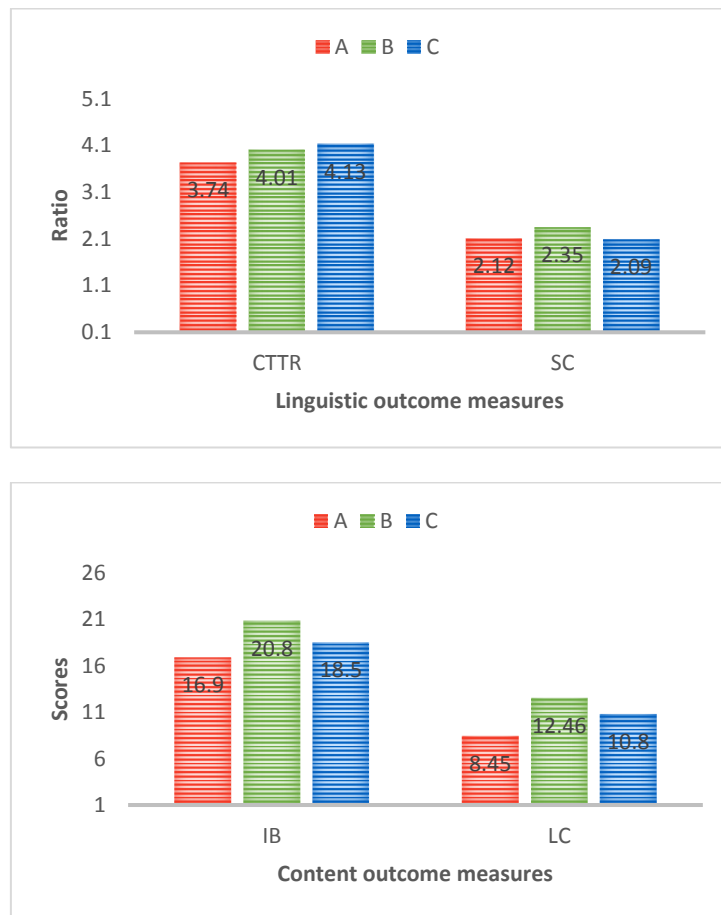
Figure 4: Differences in Story's Mean under each measure

4.3 Linear mixed-effects models

Because each speaker has several repeated data points and each Story is repeatedly contingent on several speakers, the dataset exhibited a unique dependency structure, rendering traditional statistical analyses unsuitable. LME models were

employed to analyze each outcome measure with the appropriate crossed random

variables (Baayen, Davidson, & Bates, 2008) using the *lme4* (version 1.1-15) package

in R (Bates, Mächler, Bolker, & Walker, 2014). Instead of conducting two separate

analyses – one in which data are averaged over participants (speakers or listeners; F_1)

and another in which data are averaged over items (stories; F_2) – LME models

combine these analyses (Cunnings, 2012). These models treat both subjects and items

as crossed random effects, allowing mean values for each participant and each item to

vary. In addition, LME modeling allows one to test effects and interactions of both

categorical and continuous predictors in one model (Cunnings 2012, Quené & van

den Bergh, 2008).

The main model for each measure was established by using hierarchical

(simple to complex) model selection (Bates, Kliegl, Vasishth, & Baayen, 2015).

Hierarchical model selection was used, as opposed to a backward stepwise model

(complex to simple) selection (Barr, Levy, Scheepers, & Tily, 2013), for a number of

reasons, which Bates et al. (2015) outlined in detail. In this experiment, the main

purpose of including a random-effects structure was "to obtain as powerful tests as

justified of the fixed effects" (Bates et al., 2015, p. 5). Hence, it was more rational to

leave out "variance components/correlation parameters from the model if they are not

supported by the data" (ibid, p. 5).

In creating a hierarchical model, each model starts as a basic model that only

includes a predictor and the random intercepts.[26] Then, fixed effects (conditions) are

added, one at a time, to assess cross-level interaction. The model with each new fixed

effect is tested for significance and goodness of fit according to a likelihood ratio test

(LRT). If the fixed effect's addition shows no difference, the fixed effect is dropped

---

[26] Random intercepts by item and by participant – in the case of this data, by Story and by Speaker.

from the model; if it does show a difference, it is kept. Then, in the next model, another fixed effect is added and tested. At every step, each subsequent model is compared to the preceding model using LRT. The final model includes fixed effects that show a difference and better fit the dataset. Based on the LRT, the model's goodness of fit is compared using multiple fit indices (Brown, 2006), including both Bayesian information criterion (BIC) and the Akaike Information Criteria (AIC). BIC and AIC are considered good criteria, as they discourage adding superfluous predictors that cause model overfitting (Bates, Maechler, Bolker, & Walker, 2015). The models with the lowest BIC and AIC values were chosen.

Given the complex, cross-level variables (Speaker Type, Story, and Listener Type) in this dataset, the initial models were fitted to test the significance of random intercepts. Models with different combinations of random intercepts were compared. Conceptually, neither Speaker (20 levels) nor Story (3 levels) alone qualified to be a random intercept, due to small variance. The Speaker and Story combination did qualify as an appropriate random effect; however, simply adding both Speaker and Story (e.g., (1|Speaker)+(1|Story)) to the model did not reflect the Speaker-Story dependency structure. Thus, a new variable was created that included the type of association that Speaker-Story exhibited: Speaker-Story (SS) matrix (S. Ross, personal communication, December 08, 2017). As shown in Table 5, the new variable was generated simply by recoding Story to include Speaker ID. For example, Story A was repeated 60 times (one for each listener) and linked to the same Speaker ID three times (e.g., Speaker 1 repeated Story A three times, one for each listener condition). The same method was applied to Stories B and C, for a total of 60 levels under SS (e.g., A1, A2, … A19, A20). That is, including SS $[X \sim (1 \mid SS)]$[27] as a random

---

[27] Where *X* is a predictor/outcome measure.

intercept in LME models allowed a comparison of each Story within each Speaker condition, such that each Speaker served as his or her own control in comparing his or her speech towards non-native and native listeners.

Listener	Story	Speaker	Speaker-Story (SS)	Speaker	Speaker-Story (SS)
ES	A	Speaker1	A1	Speaker2	A2
High	A	Speaker1	A1	Speaker2	A2
Low	A	Speaker1	A1	Speaker2	A2
ES	B	Speaker1	B1	Speaker2	B2
High	B	Speaker1	B1	Speaker2	B2
Low	B	Speaker1	B1	Speaker2	B2
ES	C	Speaker1	C1	Speaker2	C2
High	C	Speaker1	C1	Speaker2	C2
Low	C	Speaker1	C1	Speaker2	C2

Table 5: Speaker-Story (SS) matrix

The SS variable was included as a random intercept and compared to different combinations of random intercepts, using one predictor at a time. The model showed better fit in comparison to Speaker and Story, with much higher variance and SD. In addition, several models with different combinations of random intercepts were compared. The model with both Story[28] and SS [X ~ (1 | SS) + (1 | Story)] showed best fit overall for CTTR ($x2(4) = 12.08$, $p < 0.001$), IB ($x2(8) = 12.45$, $p < 0.001$), and LC ($x2(4) = 9.11$, $p < 0.001$). The model with both Speaker and SS [X ~ (1 | SS) + (1 | Speaker)] showed better fit for SC ($x2(4) = 6..04$, $p < 0.05$) and CVS1 ($x2(4) = 1..51$, $p < 0.001$). Because the variance was high in Story for CTTR, IB, and LC, a model with random intercepts SS and Story predicted those outcome measures best. Similarly, variance was high in Speaker for CVS1 and SC measures; therefore, the

---

[28] Notably, Story as a sole random intercept lacks power; thus, it is not independent of the SS random intercept. Story (3) in conjunction with SS (60) established the variance found in the models.

model that included SS and Speaker was the best predictor of those outcome

measures.

In the following sections, the established model with random intercepts was

used as the reference model to which models with fixed effects were compared. The

best-fitting model was identified using the goodness-of-fit test and was used to get the

results for each outcome across conditions (Listener Type, Speaker Type, and

Language Teaching). Since *lme4* does not produce p-values, the p-values were

estimated using the *lmerTest* package (Kuznetsova, Brockhoff, and Bojesen

Christensen, 2015), which is produced, along with the degree of freedom for the t-test,

based on Satterthwaite approximations. Finally, the controls for each condition were

set as ES for Listener Type, NS for Speaker Type, and WLT for Language Teaching.

4.4 Linguistic modifications
4.4.1 Lexical diversity
The CTTR dataset exhibited normal distribution, as shown by a Shapiro-Wilk

test of normality (W = 0.98737, p = 0.1075). Table 6 and Figure 5 show the

descriptive statistics for Listener Type by Story. The highest CTTR belonged to the

ES control group, followed by high- and, then, low-proficiency listeners.

LISTENER TYPE	A	B	C
ES	3.82	4.20	4.23
	(0.33)	(0.29)	(0.30)
HI	3.73	3.99	4.18
	(0.33)	(0.29)	(0.25)
LOW	3.68	3.84	3.99
	(0.30)	(0.35)	(0.33)

Table 6: CTTR overall descriptive statistics
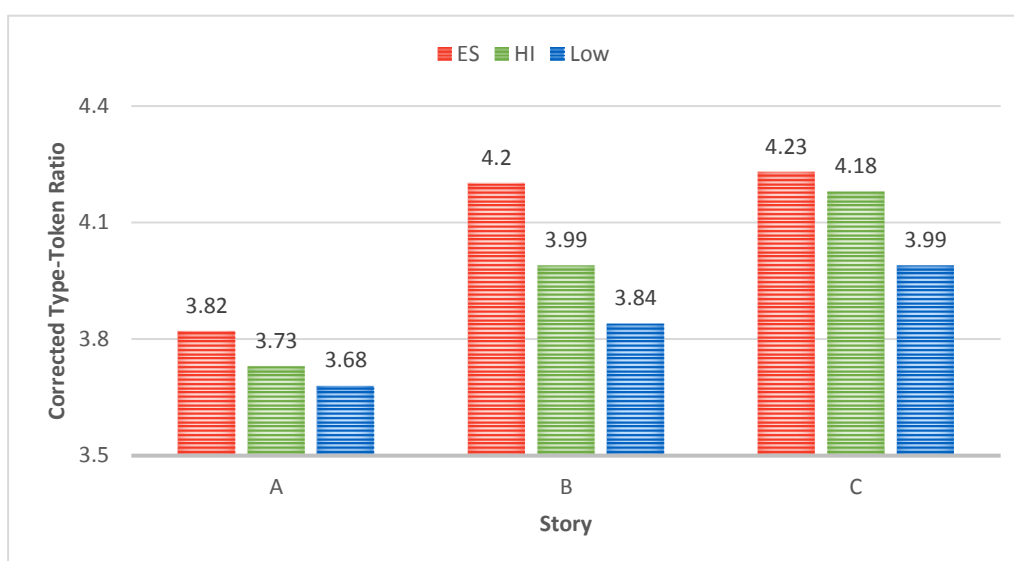
Figure 5: CTTR overall mean by Listener Type per Story

The dataset had three possible fixed effects – Listener Type (ES, Hi, and Low), Speaker Type (NS and NNS), and Language Teaching (NLT and WLT) – along with the dependent variable, CTTR. As explained earlier, the reference model only had random intercepts of SS and Story. The fixed effect Listener Type was added to the model first. The model yielded a statistical effect of Listener Type on CTTR; therefore, Listener Type was retained. Then, Speaker Type was added to the model and no effect was found, so Speaker Type was removed. The lack of difference between Speaker Types indicates that both NSs (M: 4.00 - SD: 0.32 - Range: 3.07 - 4.81) and NNSs (M: 3.93 - SD: 0.39 - Range: 2.79 - 4.74) performed similarly in their speech in the three conditions. A one-way ANOVA was conducted to further compare the effect of Speaker Types on CTTR; no significant effect was found ($F_{(1,178)}$ = 1.66, p = 0.19). Finally, Language Teaching was added to the model; it showed a significant effect on CTTR. Thus, the final model retained Listener Type and Language Teaching (LT) as fixed effects, along with SS and Story as random intercepts [CTTR~ Listener Type + LT+(1|SS)+(1|Story)].

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	4.15589	0.12158	2.51	34.182	0.0002
Listener Type [Hi]	-0.1187	0.04165	118	-2.849	0.00518
Listener Type [Low]	-0.2503	0.04165	118	-6.01	2.1E-08
LT [NLT]	-0.1328	0.06376	56	-2.083	0.04188

Table 7: CTTR overall LME model

This model was further assessed by comparing it to an interactive model (a model consists of interaction terms). The LRT indicated that the interactive model did not fit better than the main effect model (x2(9) = 2.64, p = 0.26). The results from using the main effect model showed a significant difference for Listener Type in CTTR, between ES and Hi (estimate = -0.118, SE = 0.041, t(118) = -2.84, p < 0.01) and between ES and Low (estimate = -0.250, SE = 0.041, t(118) = -6.01, p < 0.001; Table 7). Since the CTTR distribution was not transformed, the estimates in the model reflect the actual CTTR ratios. The intercept of Listener Type was ES, to which all levels in this condition were compared. Negative estimates represent ratios lower than that of ES. Figure 6 represents these results: when speaking to native listeners, speakers tended to use more lexically diverse input than when speaking to high- and low-proficiency non-native listeners, by a difference of 0.12 and 0.25, respectively.
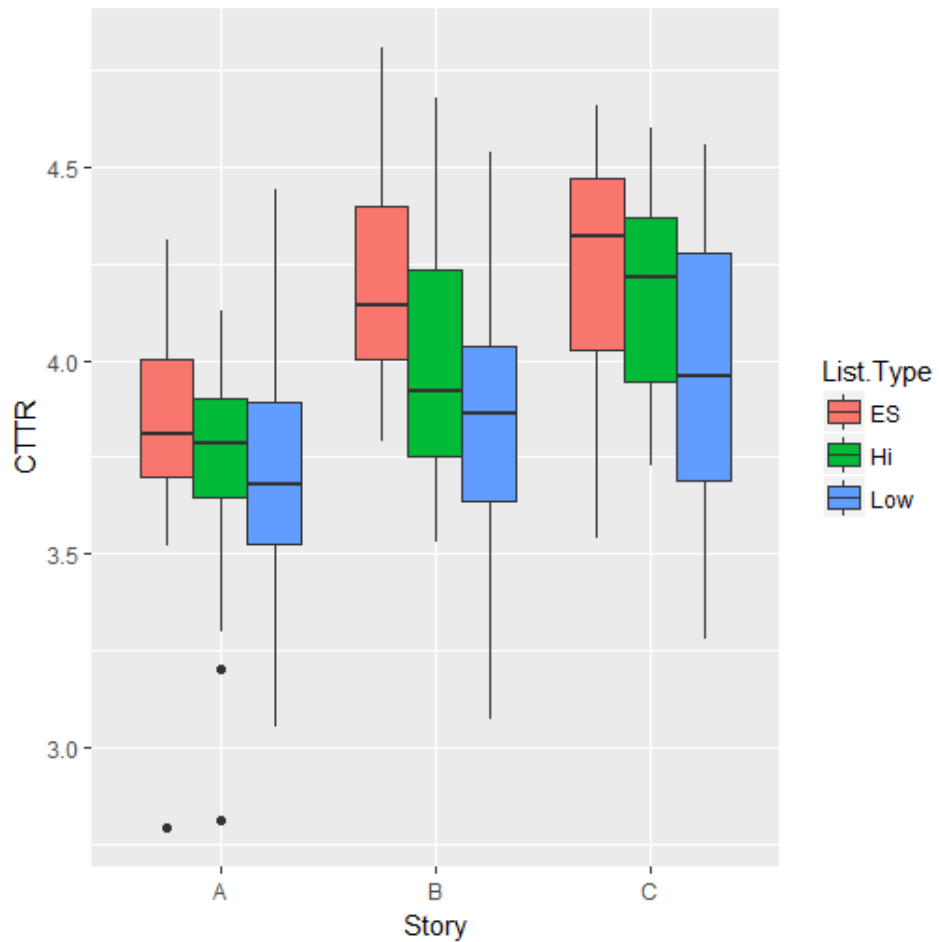
Figure 6: CTTR overall boxplot

Finally, a statistically significant difference in CTTR was found, based on Language Teaching, between WLT and NLT (estimate = -0.138, SE = 0.06, t(56) = -2.08, $p < 0.05$). These differences are depicted in Figure 7. The results imply that speakers with language teaching experience used more lexically diverse speech (M: 4.03 - SD: 0.34 - Range: 2.79 - 4.64) than speakers without language teaching experience (M: 3.90 - SD: 0.37 - Range: 3.07 − 4.81) across the three stories.
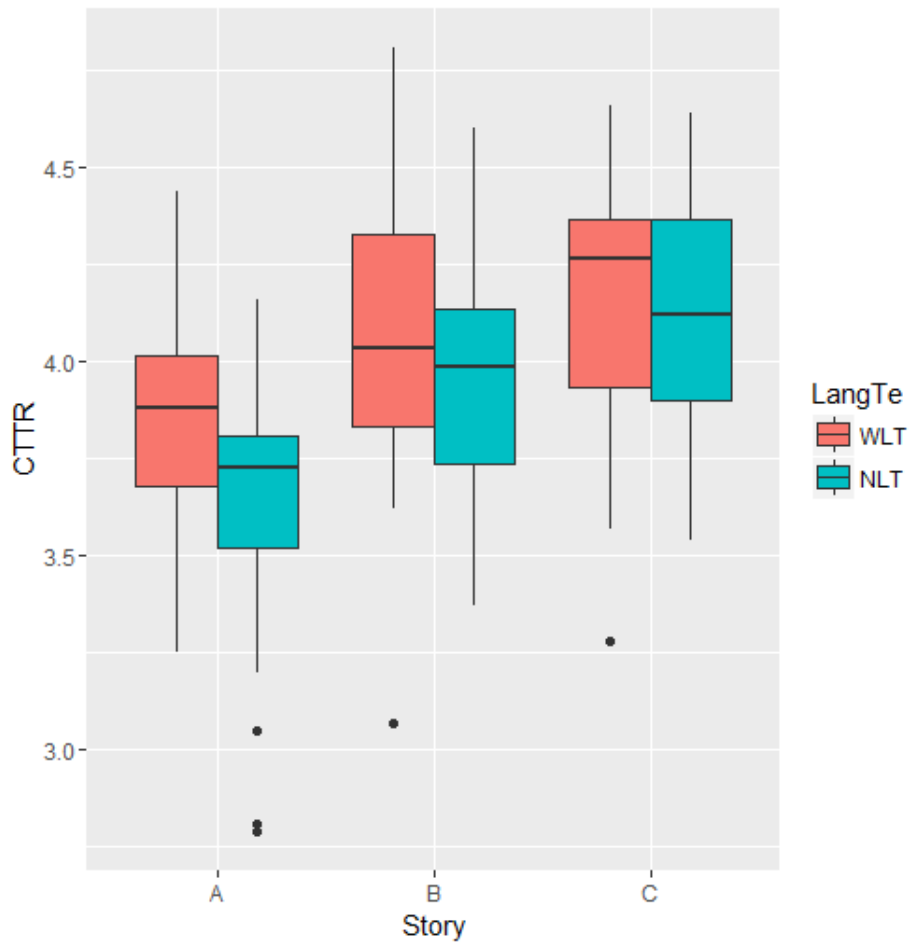
Figure 7: CTTR Language Teaching boxplot

4.4.1.1 CTTR in NSs and NNSs

The above model suggests that all Speaker Types used more repeated words (less diverse lexis) when speaking to high- and low-proficiency listeners than to native listeners. To further examine each Speaker Type's speech towards the three listener types, separate analyses of NSs' and NNSs' speech were conducted. Both analyses used the main overall model, with the exclusion of Language Teaching [CTTR~ Listener Type + (1|SS)+(1|Story)]. As indicated in Table 8, ES was set as the reference for Listener Type. The separate analysis found a significant difference in NSs' speech between ES and Hi (estimate = -0.113, SE = 0.05, t(58) = -2.24, p < 0.05) and between ES and Low (estimate = -0.300, SE = 0.05, t(58) = -5.97, p <

66

0.001). The results confirmed earlier findings that NSs use less diverse lexis in their speech towards high- and low-proficiency listeners compared to native listeners.

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	4.13933	0.0978	2.4	42.324	0.000164
Listener Type [Hi]	-0.113	0.05024	58	-2.249	0.028297
Listener Type [Low]	-0.30033	0.05024	58	-5.979	1.48E-07

Table 8: CTTR LME model (NS)

Figure 8 illustrates how NSs modified their input by reducing lexical diversity, by Listener Type and by Story. All stories followed the same pattern: CTTR was highest for ES listeners and decreased with proficiency level. Story C, *The Indian Man,* had the highest CTTR overall, followed by Story B, *The Blind Man*, and Story A, *The Elephant Weight*. While both Stories B and C dropped dramatically from ES to Low, Story A dropped only slightly from ES to Low, indicating a narrower range of lexical diversity.
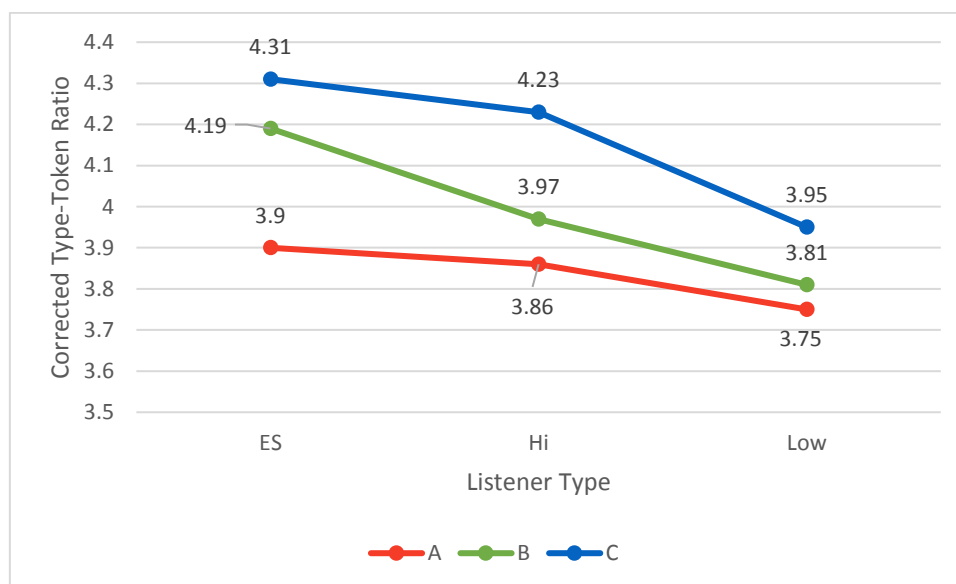
Figure 8: CTTR by listener type (NS)

NNSs' speech also confirmed a significant difference between ES and Low (estimate = -0.020, SE = 0.06, t(58) = -3.01, p < 0.01), but not between ES and Hi (estimate = -0.124, SE = 0.06, t(58) = -1.87, p = 0.06; Table 9). This suggests that

NNSs, like the NSs, modified their input to use less diverse lexis in their speech towards low-proficiency listeners.

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	4.03967	0.14618	2.31	27.635	0.000586
Listener Type [Hi]	-0.12433	0.06639	58	-1.873	0.06613
Listener Type [Low]	-0.20033	0.06639	58	-3.018	0.00378

Table 9: CTTR LME model (NNS)

As depicted in Figure 9, the modification to lexical diversity was different for each story. All stories followed the same pattern: CTTR started high for ES listeners and decreased with proficiency level. Stories C and B had the highest CTTR overall, followed by Story A. While CTTR in Stories B and C steadily decreased with listener proficiency, Story A showed only a subtle drop from ES to high-proficiency and almost no difference between high- and low-proficiency, indicating a narrower range of lexical diversity.

Overall, the results for the individual analyses show that Story A elicited the least lexically diverse speech from both NSs and NNSs. In addition, both NSs and NNSs modified their input by repeating words when speaking to non-native speakers; however, NNSs showed a smaller reduction in CTTR, possibly indicating a restricted range of vocabulary in comparison to NSs.
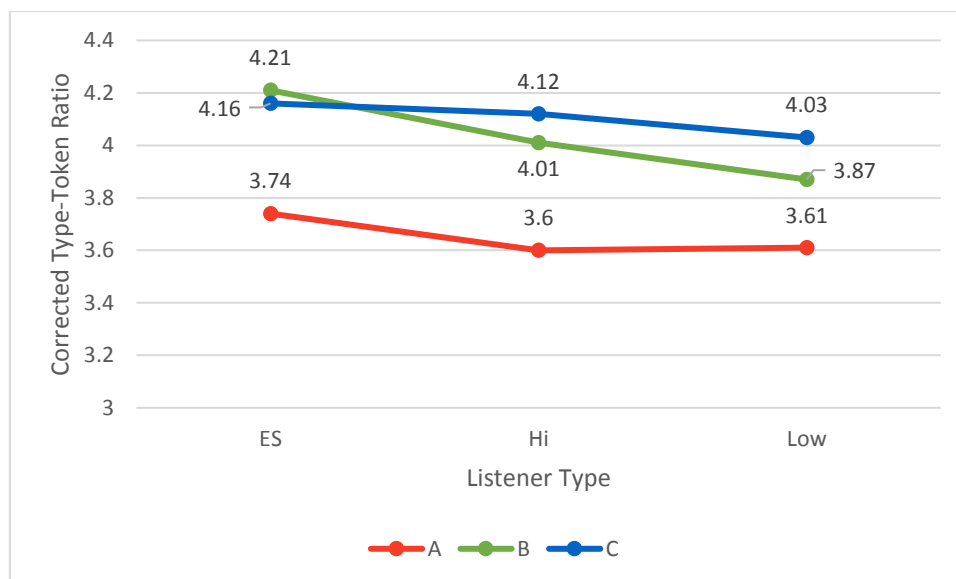
Figure 9: CTTR by Listener Type (NNS)

4.4.2 Lexical sophistication

The CVS1 ratio's distribution was positively skewed, so it was transformed via square root transformation. The Shapiro-Wilk test of normality confirmed that the transformed CVS1 distribution was normal ($W = 0.97023$, p-value $= 3.66$). Table 10 and Figure 10 show the descriptive statistics for each Listener Type under each Story. As can be seen, the highest CVS1 belonged to the ES controls, followed by high- and then low-proficiency listeners.

LISTENER TYPE	A	B	C
ES	0.339	0.317	0.328
	(0.13)	*(0.15)*	*(0.19)*
HI	0.259	0.279	0.317
	(0.11)	*(0.17)*	*(0.12)*
LOW	0.256	0.265	0.318
	(0.13)	*(0.19)*	*(0.17)*

Table 10: CVS1 overall descriptive statistics
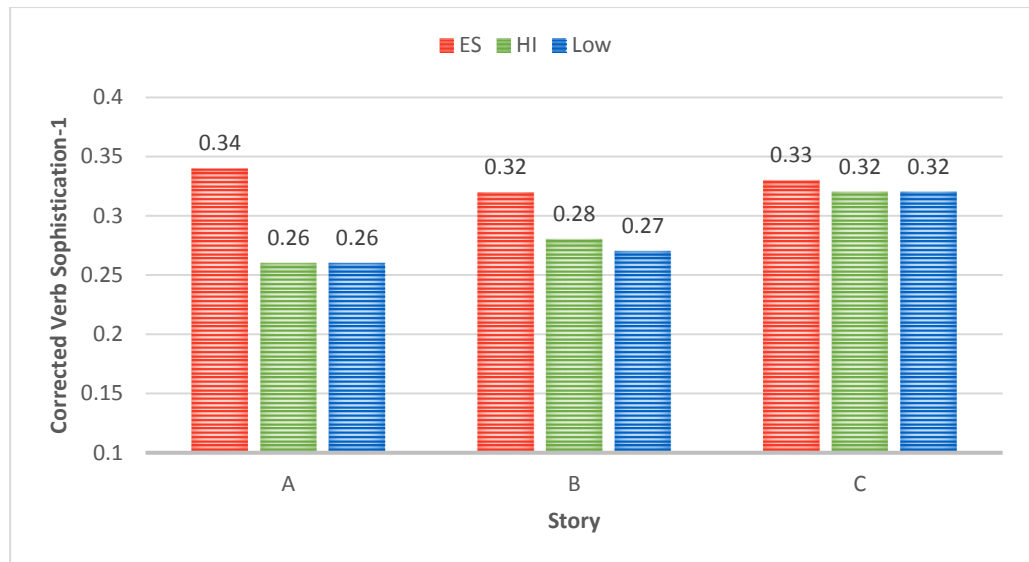
Figure 10: CVS1 overall mean by Listener Type per Story

As mentioned earlier, the reference model for CVS1 only included random intercepts for SS and Speaker. A new model was built by adding and testing one fixed effect at a time. The final model had Listener Type and Speaker Type as fixed effects and SS and Speaker as random intercepts [CVS1~ Listener Type + Speaker Type+(1|SS)+(1|Speaker)]. Unlike CTTR, there no difference was found for Language Teaching, indicating that both WLT (M: 0.31 - SD: 0.15 - Range: 0.01 - 0.84) and NLT (M: 0.28 - SD: 0.16 - Range: 0.01 - 0.76) performed similarly in their speech towards the three listener conditions. A one-way ANOVA was conducted to further compare the effect of LT on CVS1 and found no significant effect ($F_{(1,178)} = 1.384$, $p = 0.241$).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	0.615	0.02407	91.45	25.547	0.0002
Listener Type [Hi]	-0.04081	0.01949	118	-2.094	0.0384
Listener Type [Low]	-0.0479	0.01949	118	-2.457	0.0155
Speaker Type [NNS]	-0.12615	0.0301	58	-4.192	9.6E-05

Table 11: CVS1 overall LME model

70

This main model was then compared to an interactive model. The LRT indicated that the interactive model did not fit better than the main model ($x2(9)$ = 0.52, p = 0.76). The main model's results showed a significant difference for Listener Type in CVS1, between ES and Hi (estimate = -0.040, SE = 0.019, t(118) = -2.09, p < 0.05) and between ES and Low (estimate = -0.047, SE = 0.019, t(118) = -2.45, p < 0.05; Table 11). The model used ES as the reference to which the other Listener Types were compared. As depicted in Figure 11, the results indicated that speakers used the most lexically sophisticated input when narrating the story to native listeners (M: 0.328), followed by high-proficiency listeners (M: 0.285) and finally low-proficiency listeners (M: 0.280).
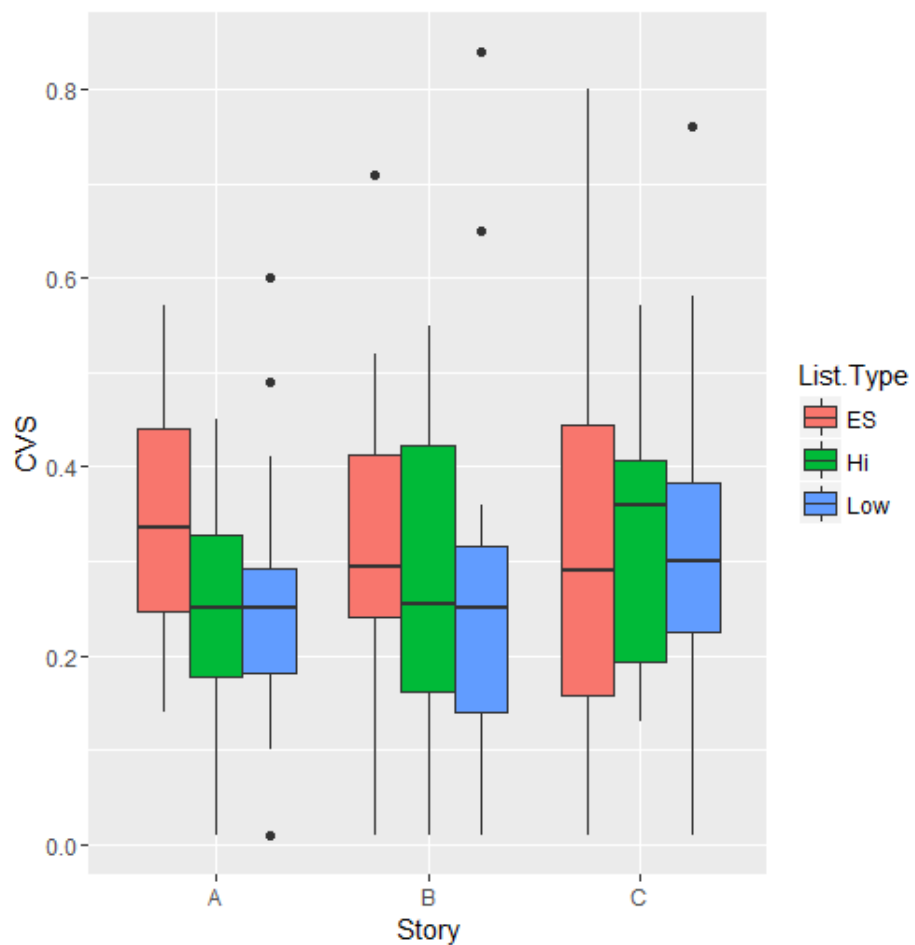
Figure 11: CVS1 overall boxplot

Story C, however, was an exception to this trend. For that story, descriptively, ES listeners were given on average *less* sophisticated input than the high- and low-proficiency listeners. This occurred perhaps due to the wide range of CVS1 for ES in Story C (0.01 - 0.8). Finally, the results also showed a statistically significant difference in CVS1 between NSs and NNSs (estimate = -0.126, SE = 0.03, t(58) = -4.19, p < 0.001), indicating that NSs used more sophisticated lexis, regardless of listener proficiency (M: 0.35 - SD: 0.15), than NNSs (M: 0.23 - SD: 0.14). Figure 12 shows this difference. While NSs displayed a wide range of lexical sophistication (Range: 0.13 - 0.84), the NNS range was more restricted (Range: 0.01 - 0.58). A possible explanation is that it may be harder for NNSs to adjust their lexical sophistication for different listener types. Separate analyses for each Speaker Type were conducted to investigate possible changes in lexical sophistication when narrating to high- and low-proficiency listeners as opposed to ES.
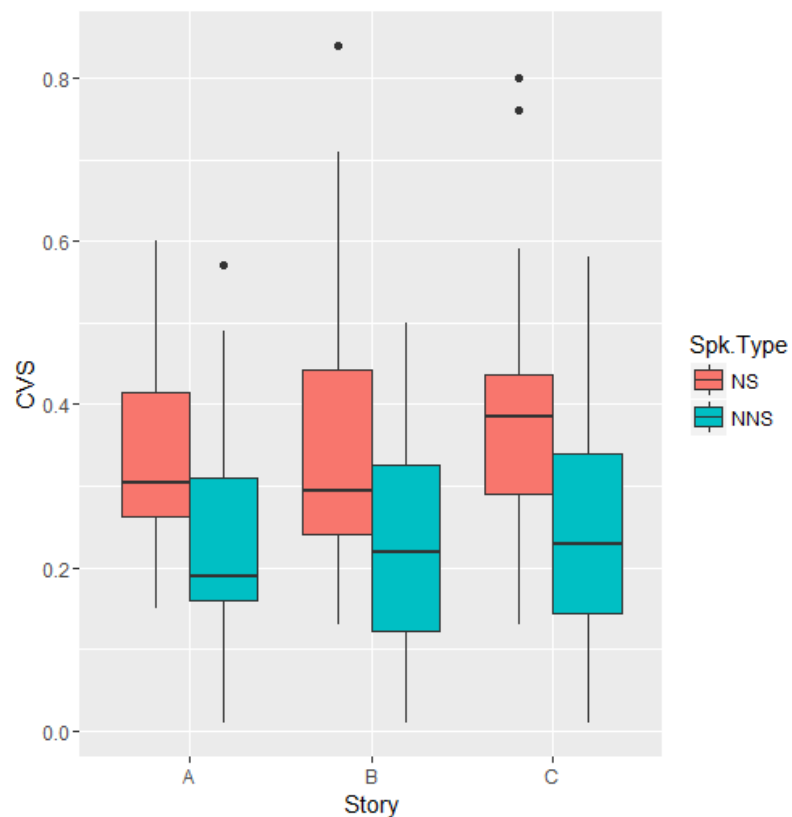
Figure 12: CVS1 Speaker Type boxplot

4.4.2.1 CVS1 in NSs and NNSs

In the following analyses, NS and NNS conditions were analyzed separately.

For NSs, the main model was used with the exclusion of Speaker Type (CVS1~

Listener Type +(1|SS)+(1|Speaker)). Analysis using this model showed no significant

difference between ES and Hi (estimate = -0.032, SE = 0.02, t(58) = -1.33, p = 0.188)

or between ES and Low (estimate = -0.034, SE = 0.02, t(58) = -1.40, p = 0.165).

However, Figure 13 and Table 12 show that the lack of effect was due to a handful of

outliers, who shifted the mean in the analysis (Range [ES: 0.13 - 0.80], [Hi: 0.13 –

0.57], [low: 0.14-0.84]). Therefore, instead of square root transformation, cut-offs for

CVS1 were established at 1.5 standard deviations above each Listener Type mean.

Transformed CVS1 greater than $Q_3 + 1.5(IQR$[29]$)$ were excluded from the analyses.

The transformation resulted in the exclusion of five data points (4.5%) of the 90 total

observations in the NSs' dataset.

LISTENER TYPE	A	B	C
ES	.381	.395	.376
	(0.11)	(0.14)	(0.20)
HI	.299	.336	.391
	(0.08)	(0.14)	(0.11)
LOW	.316	.328	.397
	(0.12)	(0.23)	(0.15)

Table 12: CVS1 descriptive statistics (NS)

---

[29] Q_3 is the third quartile, and IQR is the interquartile range. Since the lowest CVS1 score is 0.01, the "smaller than" cut-offs were not established.
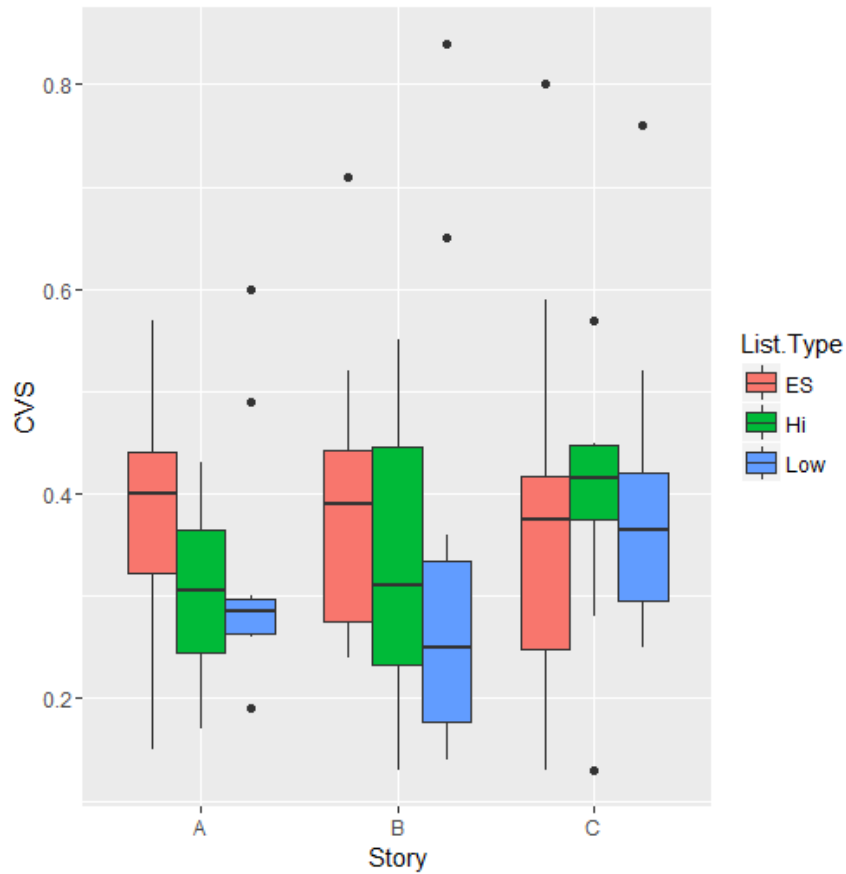
Figure 13: CVS1 boxplot (NS)

The analysis using the main model was rerun on the NS CVS1 data after excluding outliers. The new results showed a significant difference between ES and Low (estimate = -0.081, SE = 0.034, $t(87)$ = -2.36, $p < 0.05$), but no difference between ES and Hi (estimate = -0.008, SE = 0.034, $t(87)$ = 0.241, p = 0.81; Table 13). Thus, in this analysis, NSs' speech to ES (M: 0.33 - SD: 0.14) was significantly more sophisticated than their speech to low-proficiency listeners (M: 0.25 – SD: 0.13), but of similar lexical sophistication to their speech to high-proficiency listeners (M: 0.34 - SD: 0.11).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	0.333667	0.024459	87	13.642	<2e-16
Listener Type [Hi]	0.008333	0.034591	87	0.241	0.8102
Listener Type [Low]	-0.08167	0.034591	87	-2.361	0.0205

Table 13: CVS1 LME model (NS)

74

The same main model, excluding Speaker Type (CVS1~ Listener Type +(1|SS) +(1|Speaker)), was used to analyze NNSs' speech to different listener types. This model showed no significant difference between ES and Hi (estimate = -0.049, SE = 0.03, t(58) = -1.59, p = 0.115) or between ES and Low (estimate = -0.061, SE = 0.03, t(58) = -2.001, p = 0.0501), although the latter difference approached significance. After removing outliers, the difference between ES and Low in NNSs' speech was statistically significant.

The analysis using the main model was rerun on the NNSs' CVS1 data after excluding outliers: one outlier was identified and removed. The new results showed a significant difference in CVS1 between ES and Low (estimate = -0.078, SE = 0.027, $t(58) = -2.87$, $p < 0.01$), but no difference between ES and Hi (estimate = -0.043, SE = 0.027, $t(58) = -1.603$, $p = 0.11$; Table 14). In this analysis, NNSs' speech to ES (M: 0.27 - SD: 0.15) was significantly more sophisticated than their speech to low-proficiency listeners (M: 0.19 – SD: 0.11), but of similar lexical sophistication to their speech to high-proficiency listeners (M: 0.22 - SD: 0.13).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	0.27233	0.02597	21.31	10.485	7.15E-
Listener Type [Hi]	-0.04367	0.02723	58	-1.603	0.11427
Listener Type [Low]	-0.07833	0.02723	58	-2.876	0.00562

Table 14: CVS1 LME model (NNS)

Figure 14 shows little variation in means across listener conditions, but the NNSs exhibited a wide range in their speech towards ES in Story C, as opposed to low-proficiency listeners. This difference may have occurred because of individual differences, such as speaker's style. As mentioned earlier, Speaker showed a high variance in CVS1 and therefore qualified as a random intercept in the model; this variance was significant enough to cause the exhibited variance in Story C.
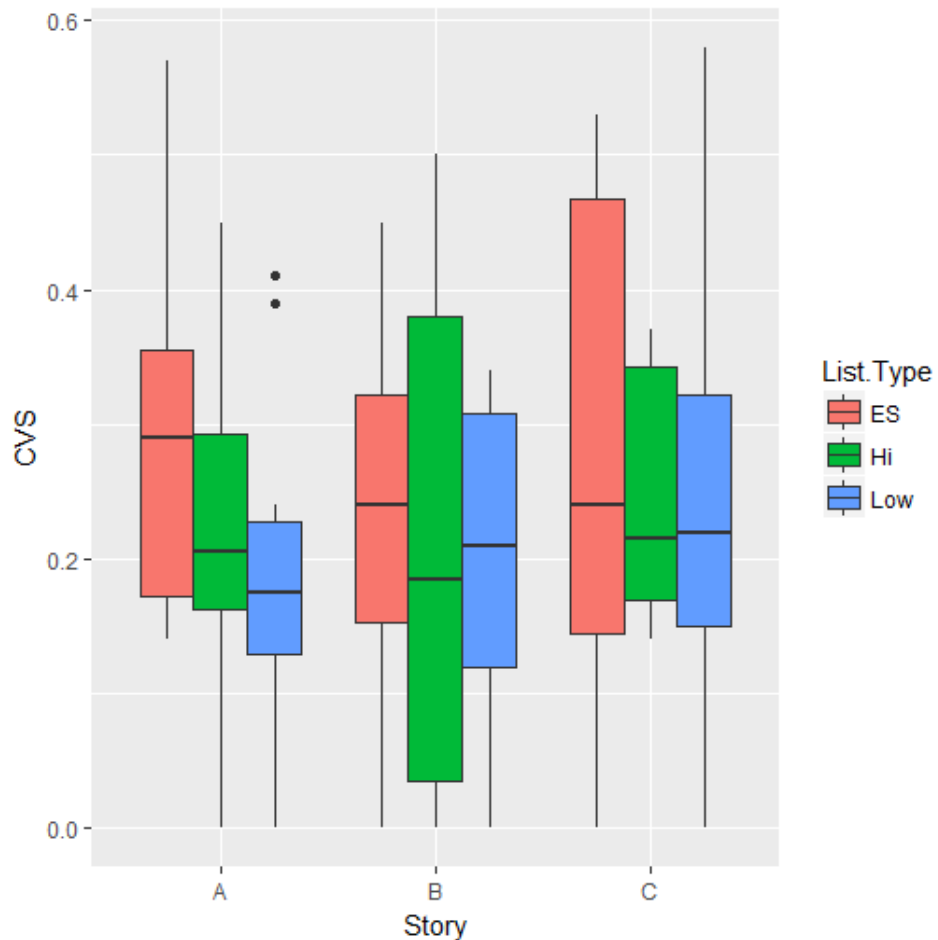
Figure 14: CVS1 boxplot (NNS)

The results of the combined analysis, with both speaker conditions, showed a significant difference between ES in comparison to Hi and Low, indicating that a possible lack of power in the separate analyses may have prevented a statistically significant difference from being detected; the separate analyses only included 90 observations as opposed to 180 in the main analysis. The overall results indicated that both NSs and NNSs tended to use less sophisticated input when speaking to low-proficiency listeners.

4.4.3 Syntactic complexity

The distribution of SC ratio was positively skewed, so it was log-transformed to reduce the skewness. The Shapiro-Wilk test of normality indicated that the log-transformed dataset was normally distributed (W = 0.97023, p-value = 0.112). Table

15 and Figure 15 show the non-transformed descriptive statistics for each listener type under each story. Based on the descriptive table and bar graph, the native controls (ES) exhibited the highest SC, followed by high- and then low-proficiency listeners.

LISTENER TYPE	A	B	C
ES	2.30	2.69	2.32
	(0.45)	(0.36)	(0.38)
HI	2.09	2.28	2.13
	(0.31)	(0.36)	(0.46)
LOW	1.98	2.06	1.82
	(0.27)	(0.38)	(0.33)

Table 15: SC overall descriptive statistics

Figure 15: SC overall mean by Listener Type per Story

The reference model for SC had random intercepts of SS and Speaker. A new model was built by adding and testing one fixed effect at a time. The resulting main model had Listener Type and Speaker Type as fixed effects, along with SS and Speaker as random intercepts [SC~ Listener Type + Speaker Type+(1|SS)+(1|Speaker)]. No difference was found between Language Teaching conditions, indicating that both WLT (M: 2.24 - SD: 0.44 - Range: 1.51 - 3.38) and

NLT (M: 2.13 - SD: 0.42 - Range: 1.42 - 3.53) used similarly complex syntax in their speech in the three listener conditions. A one-way ANOVA was conducted to further investigate the effect of LT on SC and also found no significant effect ($F(1,178) = 2.9$, $p = 0.09$).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	0.93269	0.0416	3.47	22.423	7E-05
Listener Type [Hi]	-0.11772	0.02243	118	-5.248	6.9E-07
Listener Type [Low]	-0.22031	0.02243	118	-9.82	2.00E-01
Speaker Type [NNS]	-0.10956	0.03165	56	-3.462	0.00103

Table 16: SC overall LME model

This main model was then compared to an interactive model. The LRT indicated that the interactive model did not fit better than the main model ($x2(9) = 0.12$, $p = 0.93$). The main model's results showed a significant difference in SC based on Listener Type, between ES and Hi (estimate = -0.117, SE = 0.022, $t(118) = -5.24$, $p < 0.001$) and between ES and Low (estimate = -0.220, SE = 0.022, $t(118) = -9.82$, $p < 0.001$; Table 16). The results indicated that both speaker types used more complex syntax when narrating the stories to native speakers (M: 2.44; see Figure 16) than to high-proficiency listeners (M: 2.17) or low-proficiency listeners (M: 1.95).
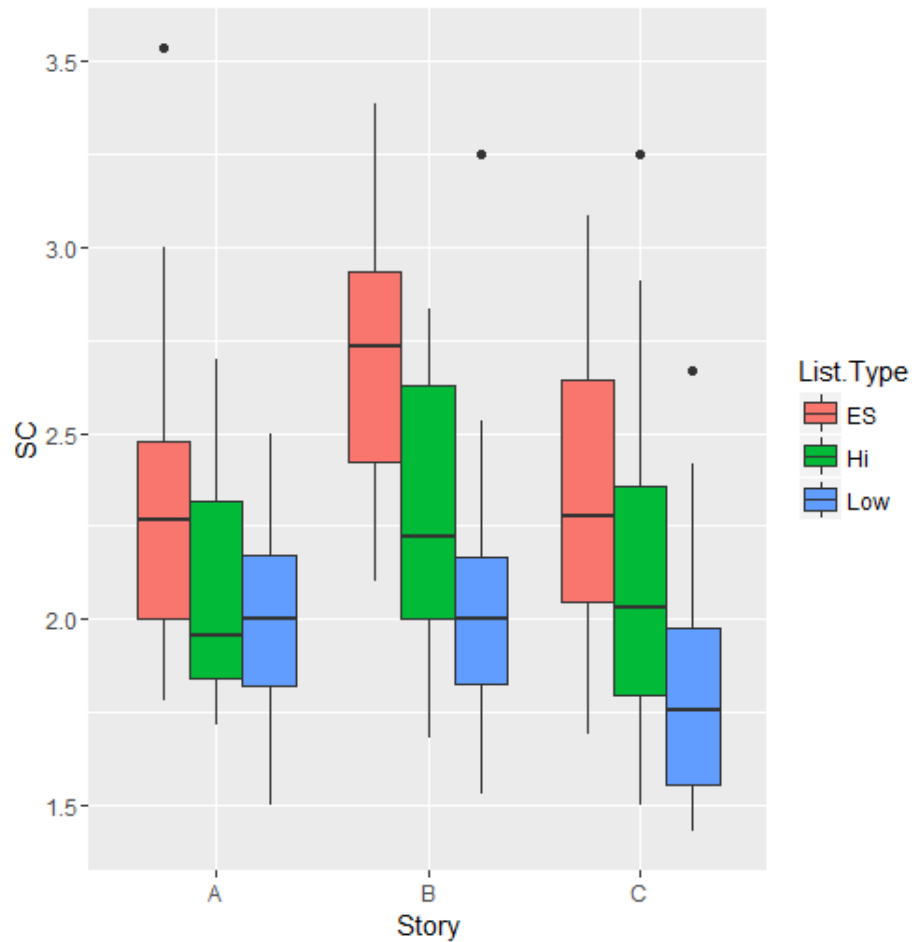
Figure 16: SC overall boxplot

Finally, the results also presented a statistically significant difference in SC for Speaker Type (estimate = -0.109, SE = 0.03, t(56) = -3.46, p < 0.001), indicating that NSs used more complex syntax in their overall speech (M: 2.31 - SD: 0.45 - Range: 1.52 – 3.53) compared to NNSs (M: 2.06 - SD: 0.38 - Range: 1.42 – 3.14). Figure 17 depicts these differences. This analysis included both speaker conditions in the model; however, to further investigate potential input modification by NSs and NNSs, separate analyses were conducted for NSs and for NNSs.
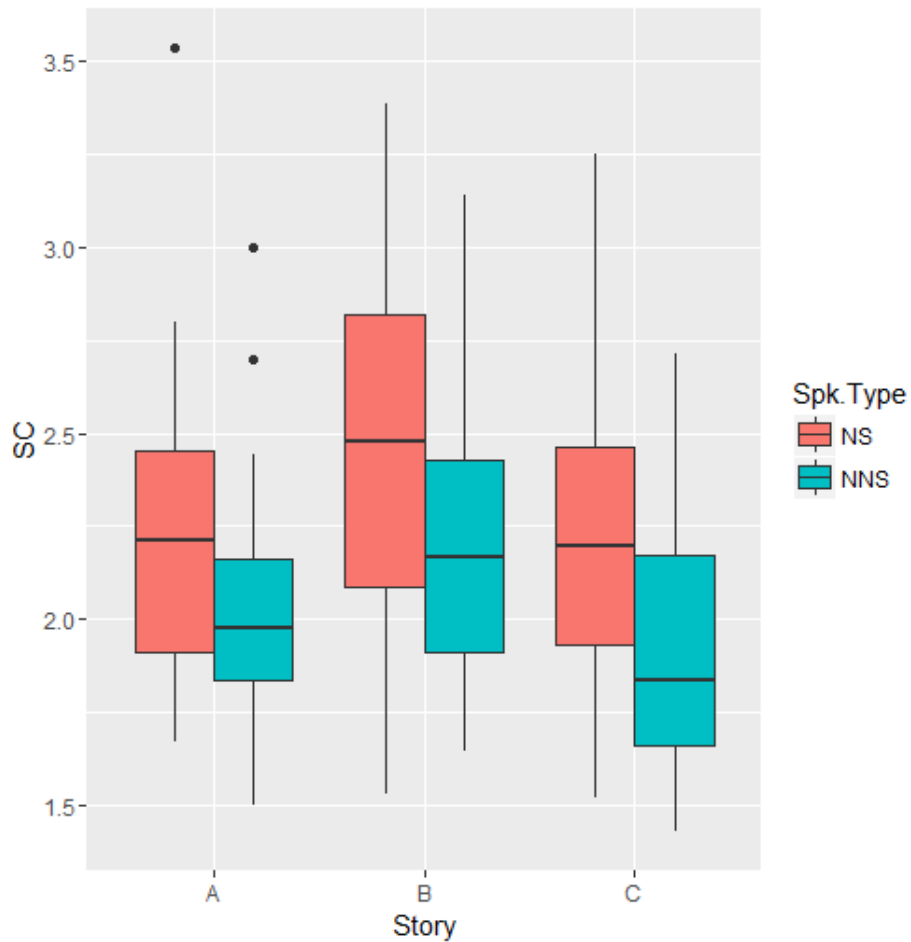
Figure 17: SC Speaker Type boxplot

4.4.3.1 SC in NSs and NNSs

In the following analyses, NSs' and NNSs' transcripts were analyzed separately. Both analyses used the main model with the exclusion of Speaker Type (SC~ Listener Type +(1|SS)+(1|Speaker)). As shown in Table 17, the separate analysis found a statistically significant difference in NSs' speech between ES and Hi (estimate = -0.116, SE = 0.03, t(58) = -3.21, p < 0.01) and between ES and Low (estimate = -0.226, SE = 0.03, t(58) = -6.25, p < 0.001). Table 18 summarizes the results of the NNS analysis, which found a significant difference between ES and Hi (estimate = -0.119, SE = 0.02, t(58) = -4.39, p < 0.001) and between ES and Low (estimate = -0.02, SE = 0.02, t(58) = -7.90, p < 0.001; Table 18). As depicted in

Figure 18, the individual results for each speaker type indicate that both NSs and

NNSs simplified their syntax according to listener proficiency.

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	0.93425	0.03163	25.98	29.54	2.00E-
Listener Type [Hi]	-0.11634	0.03622	58	-3.212	0.00215
Listener Type [Low]	-0.22638	0.03622	58	-6.25	5.25E-

Table 17: SC LME model (NS)

Fixed effects	Estimate	SE	df	T	p-value
(Intercept)	0.82157	0.038	12.96	21.621	1.49E-
Listener Type [Hi]	-0.11911	0.02709	58	-4.397	4.75E-
Listener Type [Low]	-0.21424	0.02709	58	-7.909	8.71E-

Table 18: SC LME model (NNS)
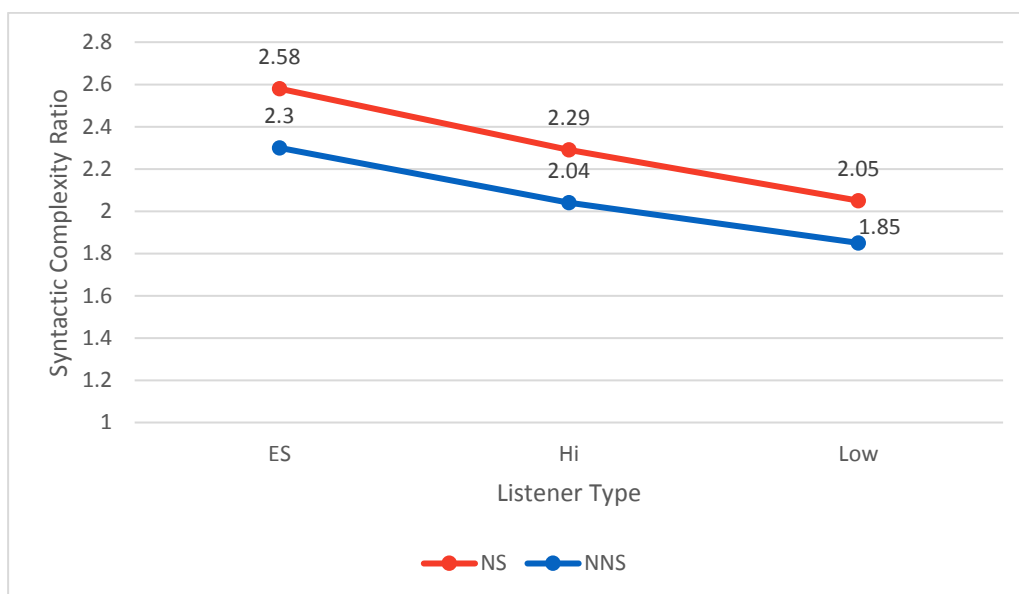
Figure 18: SC in Speaker Type by Listener Type

4.5 Content measures
4.5.1 Content dilution
Content dilution was operationalized by counting the number of mentioned

propositions, or information bits (IB), per story, and comparing the count to each

story's pre-determined list of propositions. Although the distribution of IB counts was slightly negatively skewed (W = 0.96938, p-value < 0.05), the original distribution showed better alignment to a normal distribution than any of the transformed distribution options explored, as indicated by the Q-Q plot. In addition, the Q-Q plot for IB did not show a large deviation from the diagonal line, indicating that the distribution should be considered normal. Table 19 and Figure 19 show the descriptive statistics for each listener type by story. As can be seen, the highest IB belonged to the native control, ES, followed by high- and then low-proficiency listeners. In addition, Story A had the fewest overall IBs across all conditions.

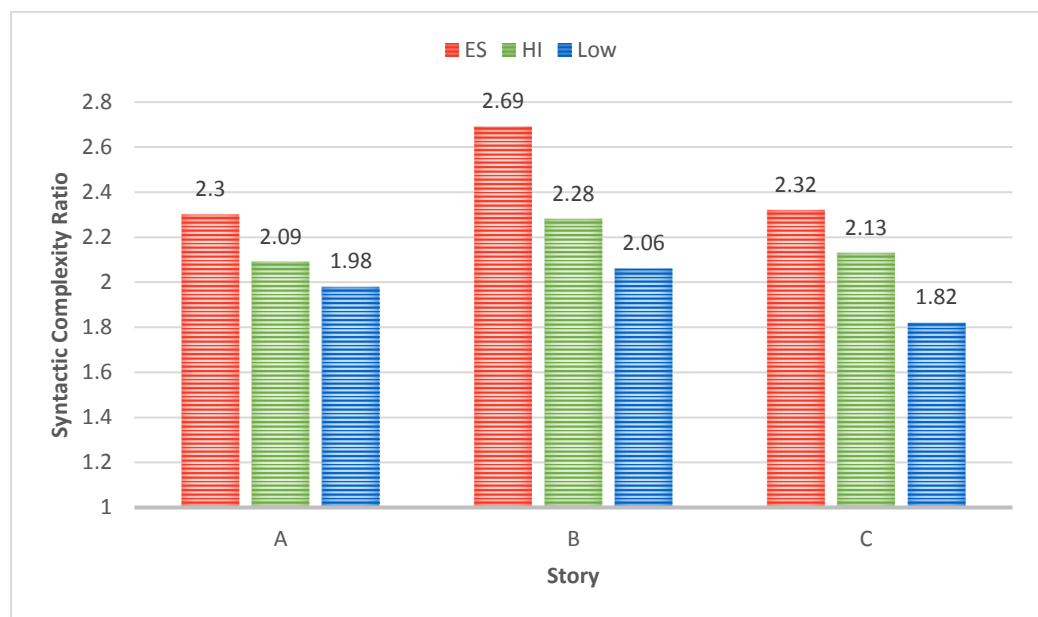| LISTENER TYPE	A	B	C
ES	18.05	21.95	18.95
	(2.72)	(1.90)	(3.17)
HI	16.80	20.40	18.90
	(2.93)	(1.66)	(3.04)
LOW	15.90	20.15	17.70
	(3.00)	(2.60)	(3.22)

Table 19: IB overall descriptive statistics
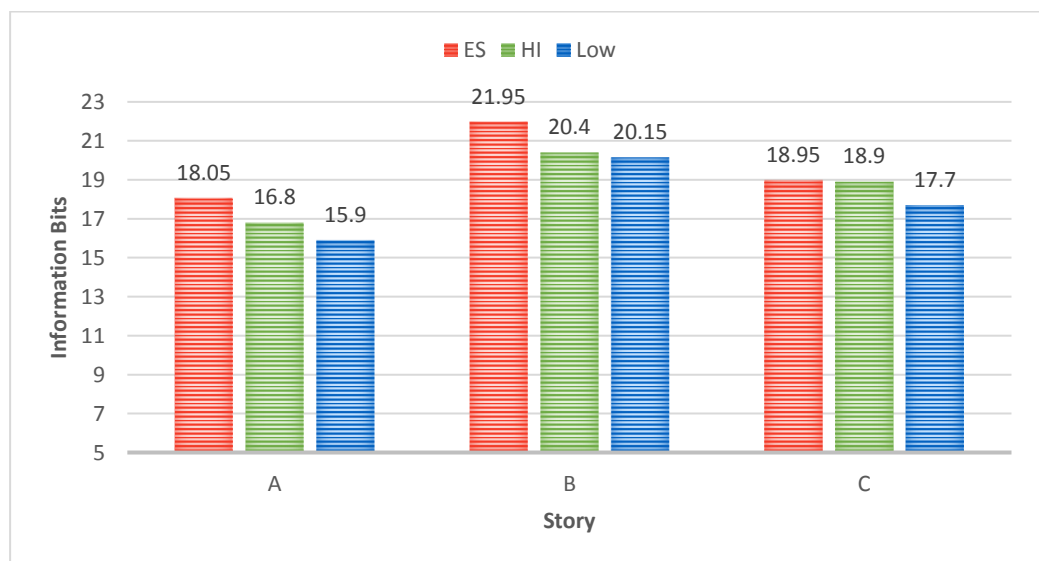
Figure 19: IB overall mean by Listener Type per Story

The reference model for IB had random intercepts of SS and Story. A new model was built by adding and testing one fixed effect at a time. The resulting main model added Listener Type as a fixed effect [IB~ Listener Type +(1|SS)+(1|Story)]. No difference was found in Speaker Type, indicating that both NSs (M: 18.86 - SD: 2.98 - Range: 13 - 25) and NNSs (M: 18.64 - SD: 3.46 - Range: 10 - 24) mentioned a similar number of IBs in their speech towards the three listener conditions. A one-way ANOVA was conducted to further compare the effect of Speaker Type on IB, and no significant effect was found ($F(1,178) = 0.21$, $p = 0.64$). Similarly, no difference was found between speakers WLT (M: 19.20 - SD: 2.95 - Range: 13 - 25) and NLT (M: 18.31 - SD: 3.43 - Range: 10 - 25), implying a similar production of content details regardless of language teaching experience. Similar to Speaker Type, a comparison of means, one-way ANOVA, was conducted to test any possible effect of LT on IB, and no significant effect was found ($F(1,178) = 3.46$, $p = 0.06$).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	19.65	1.1479	2.08	17.118	0.00286
Listener Type [Hi]	-0.95	0.2739	118	-3.468	0.00073
Listener Type [Low]	-1.7333	0.2739	118	-6.328	4.6E-09

Table 20: IB overall LME model

Because this main model only included one fixed-effect, Listener Type, a comparison to an interactive model was not applicable. Instead, the main model was compared to a model with the added fixed effects Listener Type and Speaker Type. The LRT indicated that the second model did not fit better than the main model ($x2(9) = 0.121$, $p = 0.72$). The results from the main model showed a significant difference in IB for Listener Type, between ES and Hi (estimate = -0.95, SE = 0.27, $t(118) = -3.46$, $p < 0.001$) and between ES and Low (estimate = -1.73, SE = 0.27, $t(118) = -6.32$, $p < 0.001$; Table 20), indicating that both speaker types included more IBs when narrating

the stories to native listeners (M: 19.65; see Figure 20), compared to high-proficiency

listeners (M: 18.70) and low-proficiency listeners (M: 17.91). These results suggest

that all speaker conditions increasingly excluded IBs as listener proficiency

decreased.
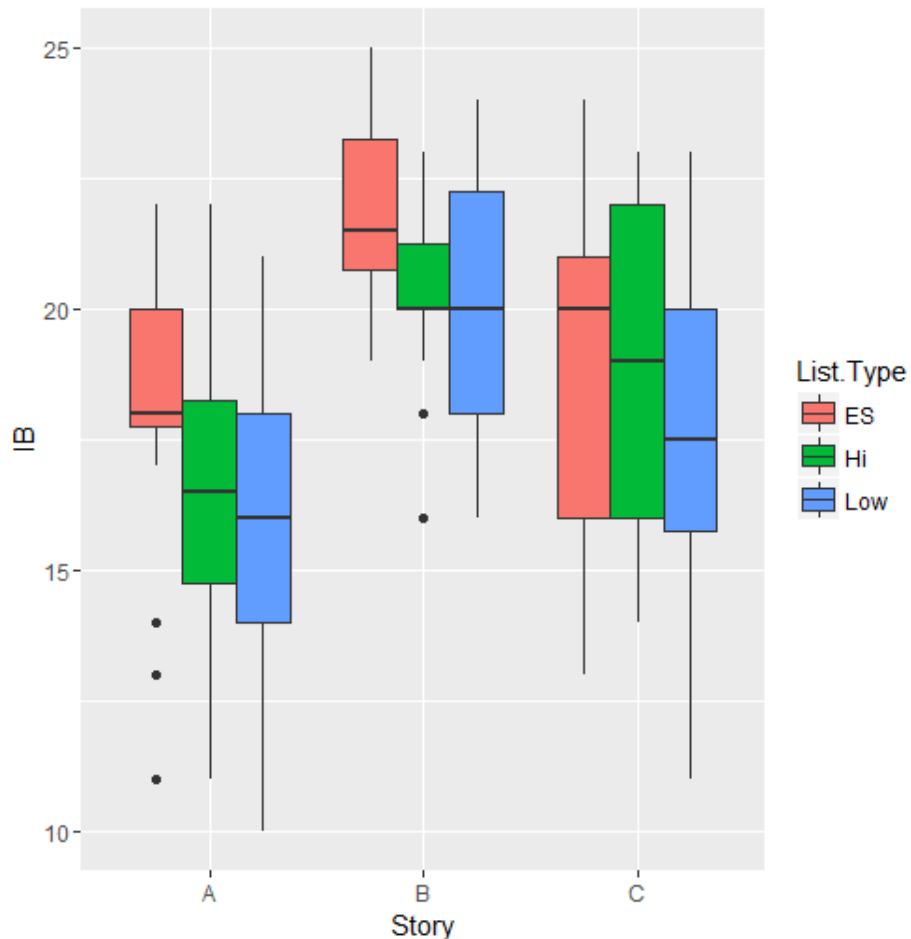
Figure 20: IB overall boxplot

4.5.1.1 IB in NSs and NNSs

To further examine the extent to which each speaker type modified their

speech towards the three types of listeners, separate analyses of NSs' and NNSs'

speech were conducted using the main model. As Table 21 shows, the separate

analyses found a statistically significant difference in NSs' speech between ES and Hi

(estimate = -1.2, SE = 0.34, t(58) = -3.46, $p < 0.001$) and between ES and Low

(estimate = -2.7, SE = 0.34, t(58) = -7.80, $p < 0.001$). The results confirmed the

findings from the main model, discussed above. Based on the estimate levels, high-proficiency listeners received, on average, at least one fewer proposition than native listeners (-1.2); low-proficiency listeners received two to three fewer propositions than native listeners (-2.7).

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	20.1667	1.1384	2.13	17.716	0.002385
Listener Type [Hi]	-1.2	0.3461	58	-3.467	0.000997
Listener Type [Low]	-2.7	0.3461	58	-7.802	1.32E-10

Table 21: IB LME model (NS)

As can be seen in Figure 21, all stories followed a similar pattern: IBs started high for ES listeners and decreased with proficiency level. Overall, Story B included the most content, followed by Story C, and then Story A. This demonstrates how speakers include different amounts of information according to content type, while overall omitting more information as listener proficiency decreases.
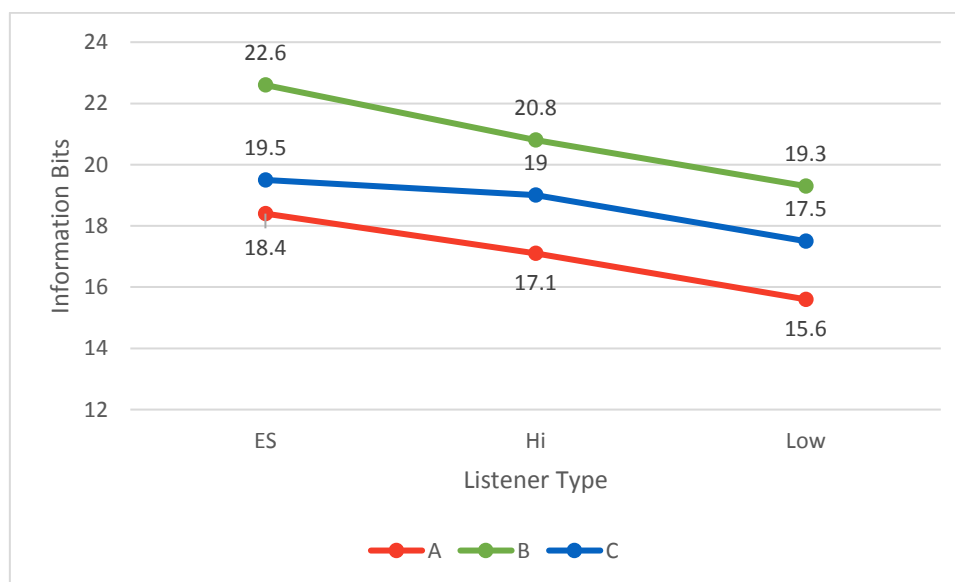
Figure 21: IB by Listener Type (NS)

As Table 22 shows, there was no significant difference in NNSs' speech between ES and Hi (estimate = -0.7, SE = 0.38, t(58) = -1.80, p = 0.07) or between ES and Low (estimate = -0.76, SE = 0.38, t(58) = -1.97, p = 0.053), although the latter

difference approached significance. Given that the initial model, which included both speaker types, showed a significant difference overall, it is possible the difference was not observed here due to lack of power – the separate analyses only included 90 observations as opposed to 180 in the main analysis. Nonetheless, descriptively, the results here implied that NNSs followed a similar pattern to that of the NSs: they modified their input by omitting more content information in their speech towards low-proficiency listeners.

Fixed effects	Estimate	SE	Df	t	p-value
(Intercept)	19.1333	1.1751	2.15	16.283	0.00271
Listener Type [Hi]	-0.7	0.3883	58	-1.803	0.07663
Listener Type [Low]	-0.7667	0.3883	58	-1.974	0.0531

Table 22: IB LME model (NNS)

Although NNSs generally reduced the content information as listeners' proficiency decreased, Figure 22 shows a different pattern for each story. Whereas Story A showed the expected continual decrease in IBs with listener proficiency, IBs for Story B decreased for high-proficiency and increased again for low-proficiency. IBs for Story C, on the other hand, increased for high-proficiency and decreased for low-proficiency. Similar to the NSs' pattern, Story B had the most included content, followed by Story C, and then Story A, which nevertheless followed a systematic decrease in mentioned IBs according to listener proficiency. This suggests that the type of content affects information inclusion. Although there was no significant difference between NSs and NNSs in IB (see p. 80), the descriptive results implies that advanced NNSs' perceptions of the content differ from NSs' perceptions. That is, the ways NSs and NNSs included and excluded content information were different across all stories.
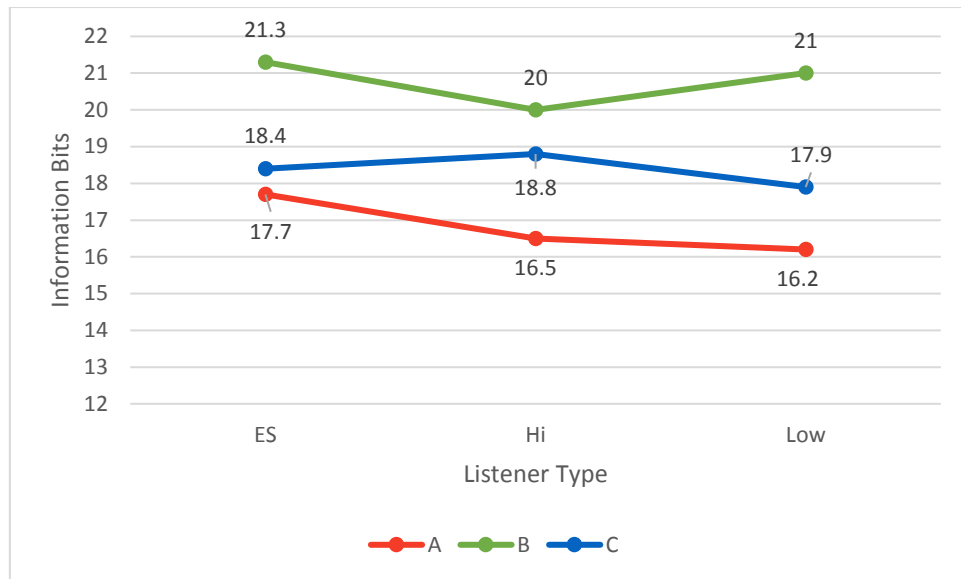
Figure 22: IB by Listener Type (NNS)

4.5.2 Listener content comprehension

All listeners' LC scores were included in the analyses (ES, High, and Low). Although the distribution of LC scores was positively skewed (W = 0.97401, p-value < 0.05), the original distribution showed better alignment to a normal distribution than the various transformation options explored, as indicated by the Q-Q plot. Table 23 and Figure 23 present the descriptive statistics for each listener type by story. As can be seen, the native controls scored the highest across stories, followed by high- and then low-proficiency listeners.

LISTENER TYPE	A	B	C
ES	11.90	17.85	13.7
	(2.88)	(3.37)	(3.84)
HI	9.60	11.95	11.2
	(5.24)	(3.53)	(3.44)
LOW	3.85	7.60	7.5
	(3.55)	(4.72)	(4.21)

Table 23: LC overall descriptive statistics

87
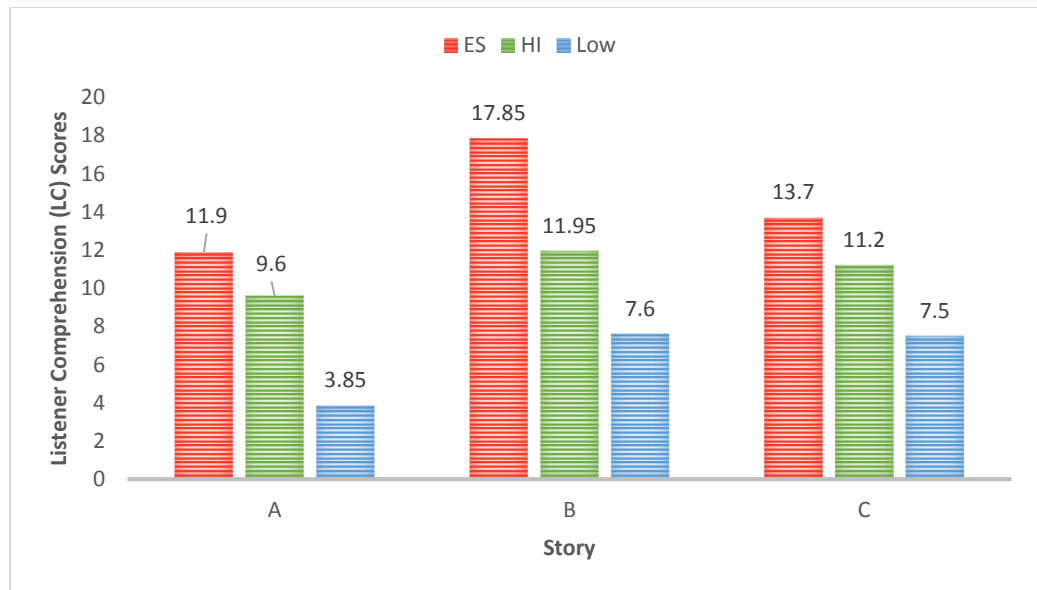
Figure 23: LC overall mean by Listener Type per Story

The reference model for LC had random intercepts of SS and Story. A new model was built by adding and testing one fixed effect at a time. The resulting main model had Listener Type and Speaker Type as fixed effects, along with SS and Story as random intercepts [LC~ Listener Type +(1|SS)+(1|Story)]. No significant difference was found between speakers WLT (M: 10.82 - SD: 5.47 - Range: 1 - 23) and NLT (M: 10.32 - SD: 5.38 - Range: 0 - 21), implying that speakers' teaching experience did not affect listeners' content comprehension.

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	15.1	1.3246	3.34	11.4	0.00087
Listener Type [Hi]	-4.6667	0.9461	116	-4.932	2.7E-06
Listener Type [Low]	-10.5	9.46E-01	116	-11.098	2.00E-1
Speaker Type [NNS]	-1.2333	0.9957	167	-1.239	0.21721
Listener Type [Hi]: Speaker Type [NNS]	2.2	1.3381	116	1.644	0.10285
Listener Type [Low]: Speaker Type [NNS]	4.6667	1.3381	116	3.488	0.00069

Table 24: LC overall LME model

This main model, which only included additive fixed-effects, was compared to an interactive model. Unlike the previous models, the LRT indicated that the interactive model did fit better than the main model ($x2(9) = 11.97$, $p < 0.05$). As shown in Table 24, the interactive model's results showed a significant difference in LC for Listener Type, between ES and Hi (estimate = -4.66, SE = 0.094, $t(116) = -4.93$, $p < 0.001$) and between ES and Low (estimate = -10.5, SE = 0.946, $t(116) = -11.09$, $p < 0.001$). Because ES was the model's baseline reference, the negative estimates for Hi and Low represent scores that are lower than the ES scores. In other words, ES listeners wrote more content details than non-native listeners across all stories (see Figure 24). Based on the estimates, high-proficiency listeners scored an average of four fewer points than ES, and low-proficiency listeners an average of 10 fewer points.
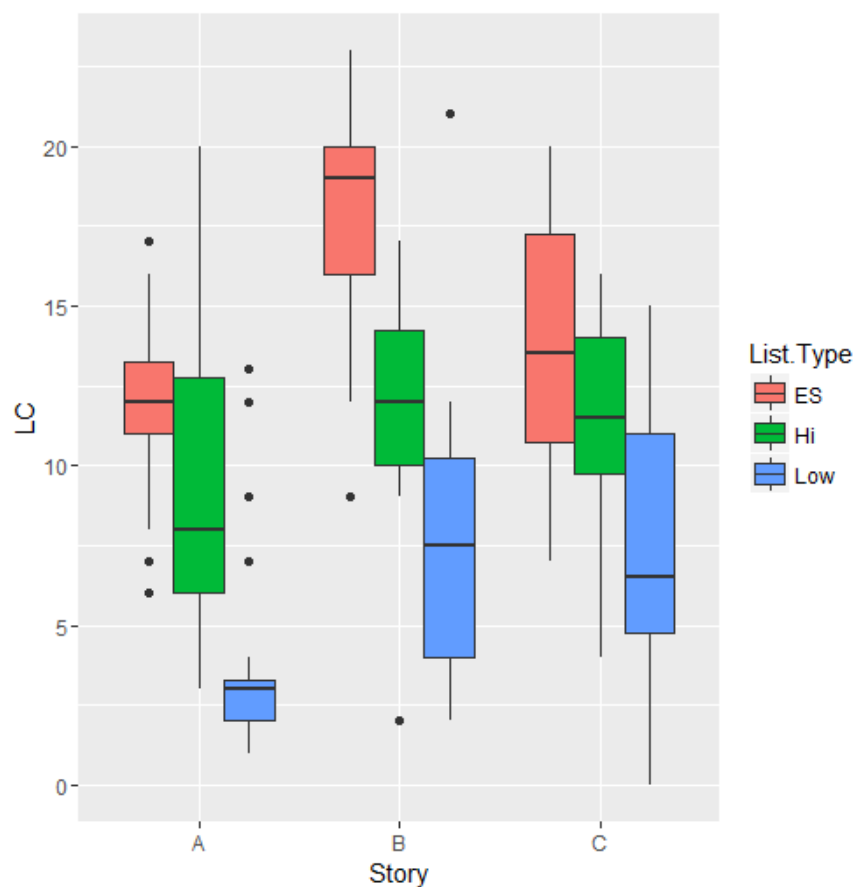
Figure 24: LC overall boxplot

Interestingly, no difference was found for Speaker Type (estimate = -1.23, SE = 0.99, t(167) = -1.23, p = 0.217), indicating that both NSs (M: 10.04 - SD: 5.76 - Range: 0 - 21) and NNSs (M: 11.10 - SD: 5.03 - Range: 1 - 23) narrated the story similarly enough that speaker type did not affect listeners' content comprehension. However, a significant interaction was found between Listener Type and Speaker Type (estimate = 4.66, SE = 1.338, t(116) = 3.48, p < 0.001). Figure 25 depicts this interaction. It shows that ES listeners who received narration from NSs scored *better* than the ES listeners who received narration from NNSs, but that the Hi and Low listeners who received narration from NSs scored *worse* than their counterparts who received narration from NNSs. In other words, native listeners scored better if they had been paired with other NSs, and non-native listeners scored better if they had been paired with other NNSs. The descriptive statistics relevant to this interaction are shown in Table 25.
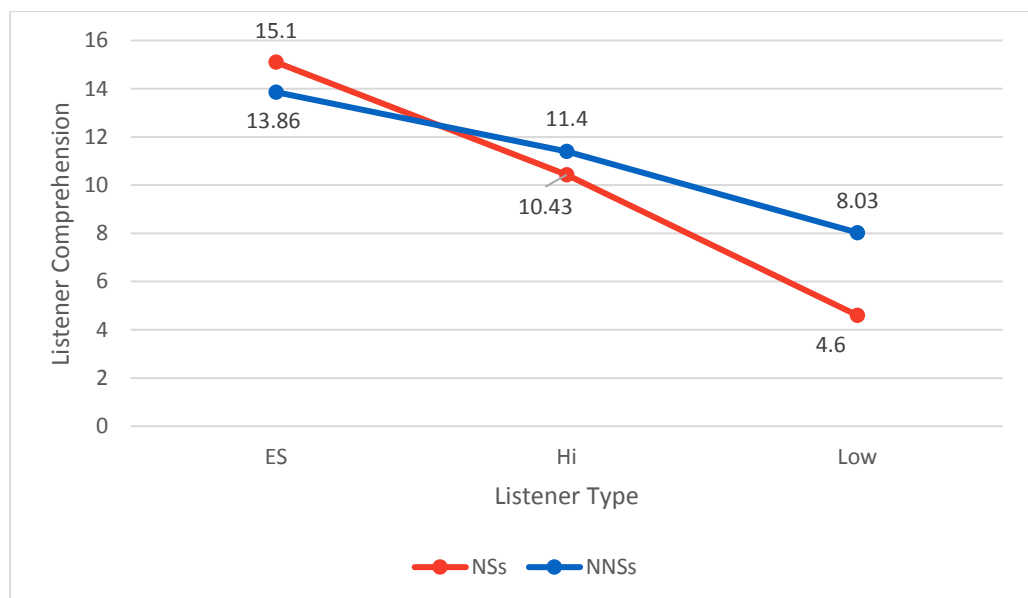
Figure 25: LC interaction between Speaker Type and Listener Type

LISTENER TYPE	NS	NNS
ES	15.1	13.8
	(3.66)	*(4.60)*
HI	10.43	11.4
	(3.11)	*(5.06)*
LOW	4.6	8.03
	(4.46)	*(3.72)*

Table 25: LC descriptive statistics for Speaker Type by Listener Type

4.5.2.1 LC between high- and low-proficiency listeners

To compare high- and low-proficiency listeners' scores, the Listener Type reference was changed to Hi, whereupon the same interactive model was used. The results exhibited a significant difference for Listener Type in LC between Hi and ES (estimate = 4.66, SE = 0.94, t(116) = 4.93, p < 0.001), and between Hi and Low (estimate = -5.83, SE = 0.946, t(116) = -6.16, p < 0.001; Table 26). The difference found between Hi and ES was the same in the previous model; however, given the positive estimate (4.66), this model indicated that ES scored, on average, four points higher than Hi, while the previous model represented this estimate in reverse (-4.66). The low-proficiency group scored significantly worse than high-proficiency listeners by five to six points (-5.83).

Finally, as presented in Table 26, no significant interaction was found between Listener Type and Speaker Type (estimate = 2.46, SE = 1.338, t(116) = 1.84, p = 0.067) using the model with Hi as the reference. This indicates that it was the low-proficiency group's difference from the native group that created the significant interaction with Speaker Type. When the reference changed to Hi, this interaction was weakened, as the difference between the groups was reduced.

Fixed effects	Estimate	SE	df	t	p-value
(Intercept)	10.4333	1.3246	3.34	7.877	0.00288
Listener Type [ES]	4.6667	0.9461	116	4.932	2.74E-06
Listener Type [Low]	-5.8333	0.9461	116	-6.165	1.05E-08
Speaker Type [NNS]	0.9667	0.9957	167.98	0.971	0.33304
Listener Type [ES]: Speaker Type [NNS]	-2.2	1.3381	116	-1.644	0.10285
Listener Type [Low]: Speaker Type [NNS]	2.4667	1.3381	116	1.843	0.06781

Table 26: LC LME: High-proficiency as the reference

4.6 Results: summary

Given the rigorous experimental design and the extent to which many

variables were controlled for, the overall results were robust for all measures. As

noted in Table 27, all speaker conditions exhibited statistically significant differences

when non-native listeners were the addressees in contrast to the native controls. Input

modification was present in lexical diversity (CTTR), lexical sophistication (CVS1),

syntactic complexity (SC), and content dilution (IB). Provided that listeners'

proficiency correlates with input modification (i.e., complexity values decrease with

the decrease of proficiency levels), it seems that proficiency levels was a cause of

input modification in general. Also, low proficiency level, resulted in low listener

content comprehension (LC) scores for both high- and low-proficiency listeners.

These differences varied depending on the story. In examining the results for each

outcome measure, it appeared that Story A yielded the lowest value for CTTR, CVS1,

and SC; it also had the fewest IBs mentioned by both speakers and listeners (in their

LC assessments), suggesting that Story A, because of its greater level of detail, was

more difficult than Stories B and C.

Linguistic and content measures in overall speech					
Listener Type	CTTR	CVS1	SC	IB	LC
ES	4.08 *(0.36)*	0.328 *(0.16)*	2.44 *(0.43)*	19.65 *(3.10)*	14.48 *(4.17)*
High	3.97* *(0.34)*	0.285* *(0.14)*	2.17* *(0.35)*	18.7* *(2.97)*	10.91 *(4.20)*
Low	3.83* *(0.34)*	0.280* *(0.16)*	1.95* *(0.34)*	17.91* *(3.39)*	6.31 *(4.48)*

* Asterisks represent statistically significant differences compared to ES

Table 27: Overall results summary

The separate analyses by Speaker Type showed similar results, with a few exceptions. As illustrated in Table 28, separate analysis of NSs confirmed the overall results but found no difference in NSs' speech between ES and Hi in lexical sophistication (CVS1). Analysis of NNSs' speech confirmed the differences found between ES and Low in CTTR, CVS1, and SC, with a significant difference between ES and Hi on SC. In addition, no differences in NNSs' speech for IB measure were found when comparing ES to Hi or Low. Although the difference between ES and Low approached significance, the lack of power in the separate analysis may have precluded a statistically significant difference.

	NSs		NNSs	
Outcome measure	ES - Hi	ES - Low	ES - Hi	ES - Low
CTTR	✓	✓		✓
CVS1		✓		✓
SC	✓	✓	✓	✓
IB	✓	✓		

Checks represent statistically significant differences

Table 28: NSs and NNSs results' comparison

Interestingly, while NSs and NNSs differed on two measures – lexical sophistication and syntactic complexity – speakers with language teaching experience

differed significantly from speakers without language teaching experience only on lexical diversity. Although NSs and NNSs followed a similar pattern overall, Story was a major factor that appeared to influence each Speaker Type's perception of the subject matter. The amount of content information that NSs and NNSs included were different across all stories, suggesting that content depth and difficulty needed to be taken into account. This difference confirmed the rationale for including three stories in the first place: to avoid the confounding variable of having only one story.

As for the LC assessment, high- and low-proficiency listeners wrote significantly fewer content details (propositions) than the native listeners, suggesting that the non-native listeners did not comprehend the stories as well as the native listeners. Importantly, based on the interaction observed in the model, while the native listeners paired with NSs condition performed better in LC than the native listeners paired with NNSs condition, the high- and low-proficiency listeners paired with the NNSs condition performed better than the high- and low-proficiency groups who were paired with the NSs condition. This suggests that NNSs mentioned more details to the Hi and Low groups than did the NSs. In the following chapter, these differences – as well as the differences between and within Speaker Type and between LT conditions – are discussed in terms of the research questions and hypotheses, with examples from the transcripts.

Chapter 5: Discussion

5.1 NSs versus advanced NNSs

Previous SLA research has studied NSs' input modifications, and results have been consistent overall. Essentially, NSs tended to simplify their input when speaking to low-proficiency listeners. However, there were no studies that investigated whether advanced NNSs also modified their input in other ways. Before discussing the input modifications exhibited in this study by both NSs and NNSs, the first research question (RQ1) – which compared NSs' and NNSs' input through four measures: lexical diversity, lexical sophistication, syntactic complexity, and content dilution – is addressed below.

5.1.1 Lexical diversity

The first hypothesis under RQ1 (H1a) stated, *NSs will use more lexically diverse speech (i.e., fewer word repetitions) than NNSs in the three listener conditions. This difference will be indicated by a higher CTTR value for NSs.* The results of the main model with both Speaker Types indicated that there were no differences in CTTR between NSs (*M: 4.00 - SD: 0.32*) and NNSs (*M: 3.91 - SD: 0.34*) speech to all listener types: ES, Hi, and Low. Descriptively, however, NSs did use more diverse words than NNSs with ES and high-proficiency listeners, indicating a non-significant difference between NSs and NNSs in lexical diversity. As exhibited in Figure 26, even though NSs means were higher, both NSs and NNSs showed a similar pattern, in which they tended to repeat more words when speaking to non-native listeners of both proficiency levels.
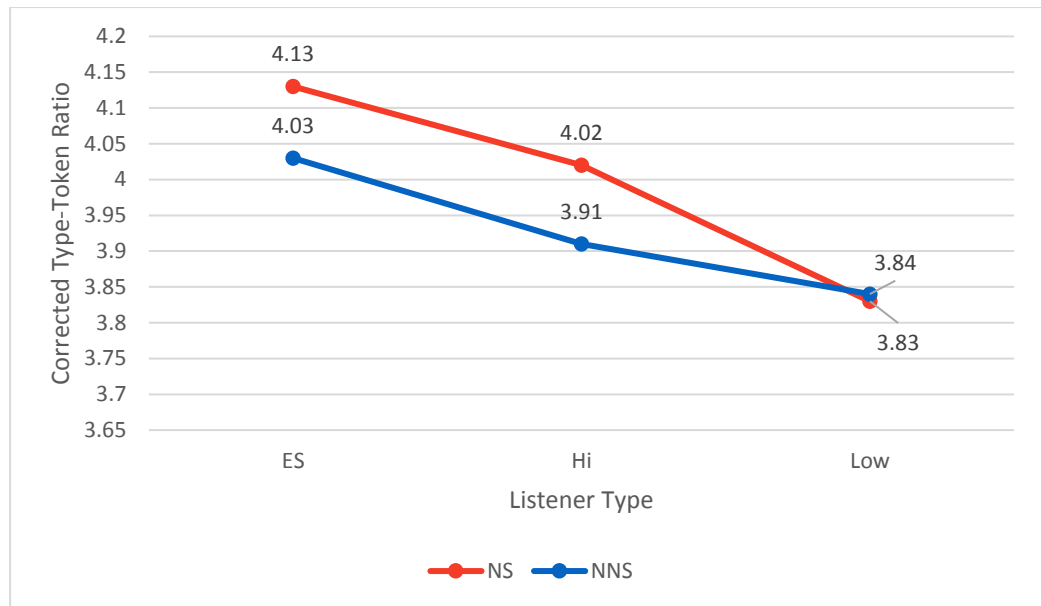
Figure 26: CTTR in Speaker Type by Listener Type

For example, as part of *The Indian Man* (Story C), both speaker types were inclined to use different vocabulary referring to similar meanings of the word "basket" in their speech to ES: "basket," "suitcase," and "box." In their speech to high- or low-proficiency listeners, however, both types of speakers repeatedly used the same word ("basket" or "box") throughout their narratives. Similarly, in *The Blind Man* (Story B), speakers used various words, such as "goes," "walks," and "approaches," to describe the scene where *the boy was going to the blind man* in their narrative to ESs; however, they tended to stick with one or two simple words, such as "goes," when narrating to non-native listeners. The results implied that both NSs and NNSs used a broader range of words when speaking to ESs, perhaps because they assumed that native speakers would comprehend more vocabulary and so they did not have to adhere to only one or two words. On the other hand, when NSs and NNSs were addressing non-native interlocutors, especially of low proficiency, their range of vocabulary became narrower; speakers may have assumed that non-native speakers would not comprehend a great variety of words, so they used the same words repeatedly throughout their narration. A similar *range* of vocabulary, however, does

96

not mean that both NSs and NNSs used the same *type* of vocabulary, as will be

explained in the following section on lexical sophistication.

5.1.2 Lexical sophistication

The second hypothesis under RQ1 (H1b) stated, *NSs will use more*

lexically sophisticated speech (i.e., more low-frequency words) than NNSs in the three

listener conditions. This difference will be indicated by a higher CVS1 value for NSs.

The main model found a statistically significant difference in CVS1 between NSs'

and NNSs' speech. Specifically, NSs used more sophisticated lexis (*M: 0.35 - SD:*

0.15) in comparison to NNSs (*M: 0.23 - SD: 0.14*). NSs, of course, have larger

lexicons than NNSs; therefore, they were more likely to incorporate low-frequency

words in their input to listeners, especially to native listeners. The NNSs,

understandably, have smaller lexicons, leading to an increased use of high-frequency

words.

NSs, for instance, used words such as "sneaks," "drops," "slams," and

"shouts," while NNSs used words such as "goes behind," "puts," "closes," and "tells,"

in the same stories. This pattern held true for nouns and adjectives, as well. NSs used

words such as "allowance," "shop," "clever," and "wealthy," while NNSs used more

common words, such as "money," "store," "smart," and "rich." Of course, individual

speakers used these terms differently – a number of NNSs used words similar to those

mentioned by NSs – but the overall difference between NSs and NNSs in lexical

sophistication was statistically significant. The following are two excerpts from the

transcripts of a NS and a NNS – each of whom was speaking to a native listener and

neither of whom had language teaching experience – in the same lines of Story A, *The*

Elephant Weight.

Excerpt A: CVS1 in speaker type (NS)

NS5 (NLT): speech to ES listener (Story A)

S: and then a little boy (a) <u>comes along</u> with an idea . that .. they can put the

elephant in a barge …. so they put the elephant in a boat .. and when that

happens . the boat sinks a little bit . so the little boy . (b) <u>paints a line</u> . to

where the (c) <u>water level</u> is ... and then they take the elephant out of the boat

and they start to (d) <u>pour rocks</u> . into the boat . until it sinks back . to the line

that the little (e) <u>boy drew</u> ... and then they take and they weigh each bucket of

rocks . one at a time and add that together to (f) <u>figure out</u> how much they

elephant weighed…

[*lines: 7 – 18*]

Excerpt B: CVS1 in speaker type (NNS)

NNS5 (NLT) speech to ES listener (Story A)

S: so . a smart boy (a) <u>came with</u> an idea of using a barge.. and he suggested

that they put the elephant on the barge .. and then once it sinks in the water .

(b) <u>he marked </u>where the (c) <u>water was</u> .. then he took out the elephant .. and

then decided that they can (d) <u>put stones</u> and as much- they put stones and as

much as they need to reach that mark (e) <u>he made</u> .. then they took out the

stones . weighed them . and (f) <u>he calculated</u> the weight of the elephant .

[*lines: 4 – 12*]

In both excerpts, the different uses of similar vocabulary are underlined. Even

though both the NS and the NNS were speaking to an ES, they used different types of

words. While the NS used "(b) he paints" and "(c) water level," for example, the NNS

used "(b) he marked" and "(c) where the water was." The NNS also used "(d) put

stones" and "(b) marked," while the NS used "(d) pour rocks" and "(b) paints,"

suggesting a greater level of sophistication by the NS. These excerpts illustrate the

differences in lexical sophistication between NSs and NNSs within the same story.

Both speaker types, however, tended to modify their input by using less sophisticated lexis when speaking to low-proficiency listeners. For example, a NS used "they *get* into a car" in his speech to a low-proficiency listener, but used "they *jump* into a car" in his speech to a native listener. These differences are explored in details in sections 5.3.2 and 5.4.2.

5.1.3 Syntactic complexity

The third hypothesis under RQ1 (H1c) stated, *NSs' speech will exhibit greater syntactic complexity than NNSs' speech in the three listener conditions. This difference will be indicated by a higher number of S-nodes per clausal utterance for NSs.* The main model showed a significant difference in SC between NSs and NNSs, with NSs using overall more syntactically complex input (*M: 2.31 - SD: 0.45*) than NNSs (*M: 2.06 - SD: 0.38*). Similar to the lexical sophistication results, those for syntactic complexity reflect NSs' superior command of the language and their ability, therefore, to incorporate more syntactically complex utterances in their input. Even though the NNSs were advanced, they still had a lower command of syntax in comparison to NSs, leading to the use of simpler utterances.

NNSs' generally lower command of syntax in comparison to NSs' has been discussed widely in the SLA literature. Scholars such as Clahsen and Felser (2006) have explained, from a psycholinguistic perspective, that grammar acquisition by NNSs, specifically late L2 learners, is generally less successful and exhibits a fundamental difference from the grammatical system of NSs. Given that all NNSs in this study studied English as adults,[30] scholars supporting the *Critical Period Hypothesis* would posit that these late adult L2 learners could not have achieved a

---

[30] They studied English during their secondary education, but were not fully exposed to daily practice and use of English until they came to the United States as adults (either for education or for work).

native-like fluency in morphosyntax (Johnson and Newport, 1989; Johnson, 1992; Schachter, 1996; DeKeyser, 2000, 2005).

In this study, NSs used a wide range of syntax (Range: 1.52 – 3.53) that differed significantly from NNSs' (Range: 1.42 – 3.14) and changed based on listener type. Speakers varied individually within their groups – a number of NNSs used utterances that were similarly complex to those employed by NSs – but the general difference in syntactic complexity between speaker types was statistically significant. The following are two excerpts from the transcripts of a NS and a NNS – each of whom was speaking to a native listener and both of whom had language teaching experience – in the same lines of Story B, *The Blind Man*.

Excerpt C: SC in speaker type (NS)

NS10 (WLT): speech to ES listener (Story B)[31]

> **S:** /... he :: looks across the street though . and he :: realizes that there :: is a blind man . there :: begging / (4/1)[32]... /the boy :: decides . *because* he :: has pangs of conscience . that he :: should donate his money to the blind man/ (3/1) ... /so he :: starts crossing the street/ (1/1) . /and *while* he :: is doing this a woman :: arrives/ (2/1). /and she :: gets out of a car . and . :: slams the door just as the boy :: is putting his coins into the blind man's cup/ (3/1)
>
> [*lines: 4- 12*]

Excerpt D: SC in speaker type (NNS)

NNS8 (WLT) speech to ES listener (Story B)

> **S:** /... suddenly he :: saw a blind man/ (1/1) . . /so out sense of guilt he :: felt like he :: should give the money to the blind man . instead of buying his- . the

---

[31] Slashes were used to mark units (clausal utterances) and double colon is used to mark each S-node.
[32] (S-nodes/clausal utterance)

toy that he[33] wants- he :: wanted to/ (3/1) ... /so he . :: put the money on . the

blind's man … cup whatever so/ (1/1) ... /and then at that same time there ::

was a lady :: coming from her car and she- she :: slammed the door/ (3/1)

[*line: 5 – 14*]

Both excerpts have the same content, and both were directed to native

listeners. Although both speakers had comparable language teaching experience, their

speech to ES exhibited syntactically different input. Whereas the NS's excerpt (C)

included a total of thirteen S-nodes and five clausal utterances (SC ratio: 2.6), the

NNS's excerpt (D) included eight S-nodes and four clausal utterances (SC ratio: 2).

The NS used more syntactically complex utterances overall, ranging from one to four

S-nodes per clausal utterance. The NNS used less complex syntax overall, ranging

from one to three S-nodes per clausal utterance. In addition, the NS appropriately

used subordinate conjunctions such as "while" and "because," thereby including an

independent clause along with one or more dependent clauses in one unit.

The NNS, on the other hand, mostly used shorter utterances (e.g., in excerpt

D, two out four utterances were simple) and tended to use coordinating conjunctions

more than subordinate ones. The repeated use of coordinating conjunctions led, in

many cases, to sentence fragments. Certainly, this was not only the case for NNSs;

NSs' speech also presented the use of fragments. Although NSs and NNSs used

syntax differently, both speaker types tended to modify their syntactic input by using

shorter utterances when speaking to low-proficiency listeners. These differences are

explored in detail in sections 5.3.3 and 5.4.3.

---

[33] Self-corrected or repeated words were not counted.

5.1.4 Content dilution

Finally, the last hypothesis listed under RQ1 (H1d) stated, *NSs will include more content details in their speech than NNSs in the three listener conditions. This difference will be indicated by higher counts of mentioned information bits (IBs) by NSs.* The main model showed no significant difference between NSs' (*M: 18.86 - SD: 2.98*) and NNSs' (*M: 18.64 - SD: 3.46*) speech. As presented in Figure 27, both speaker types showed a similar pattern in their speech towards listeners, but a distinctive pattern in relation to each other, indicating a possible interaction between Listener Type and Speaker Type.
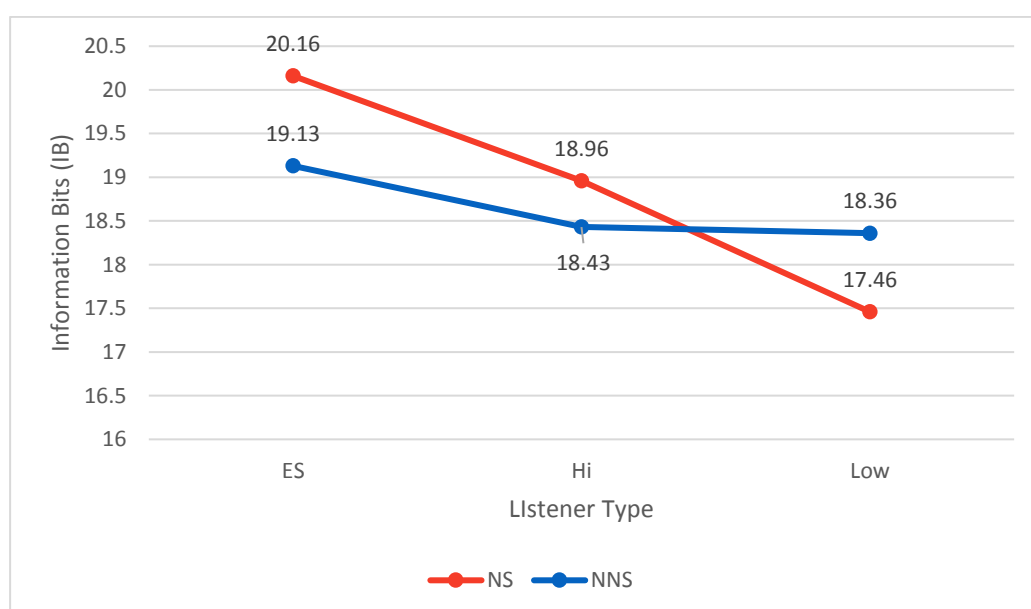
Figure 27: IB in Speaker Type per Listener Type

Although the model used in establishing results for content dilution did not show any interaction effect,[34] the interaction of Listener Type and Speaker Type was tested here in a different model that included both as fixed effects along with interaction terms [IB~ Listener Type * Speaker Type*(1|SS)*(1|Story)]. Interestingly, a significant interaction was found between Listener Type and Speaker Type

---

[34] Because the best fitting model did not include Speaker Type as a fixed effect in the main model formula (see section 4.5.1 for details)

(estimate = 1.93, SE = 0.52, t(116) = 3.71, p < 0.001), indicating that in comparison to NNSs, NSs included *more* content information in their speech to native and high-proficiency listeners, but *less* content information in their speech to low-proficiency listeners. This interaction resulted because NSs' mean information bits continually decreased from ES to Low (ES: 20.16, Hi: 18.96, and Low: 17.46), while NNSs' mean of information bits decreased from ES to Hi but remained almost the same between Hi and Low (ES: 19.13, Hi: 18.43, Low: 18.36). This indicates that NNSs omitted a similar number of information bits when speaking to high- and low-proficiency listeners. NSs excluded a greater number of information bits in their speech to low-proficiency listeners than when speaking to the native listeners.

The interaction effect found in IB also reflects the interaction effect found in LC. As mentioned in the results section, high- and low-proficiency listeners who were paired with NNSs outperformed high- and low-proficiency listeners paired with NSs, suggesting that NNSs retained more information bits in their speech to the non-native listeners. It is unclear what made NNSs retain more information compared to NSs in their narration to non-native listener; however, investigating these observations is vital to future research on the subject.

5.1.5 Summary

This study demonstrated some of the differences between NSs and NNSs discussed in the SLA literature. For instance, it showed that NSs differed statistically significantly from NNSs in the lexical sophistication and syntactic complexity of their speech. However, there was no statistical difference found between NSs and NNSs in lexical diversity and content dilution, indicating that both speaker types performed similarly in these two measures. Table 29 summarizes these results. This study also showed an interaction effect between Speaker Type and Listener Type on both the listener content comprehension and the mentioned information bits, suggesting a

deeper level of difference between Speaker Type that could not be explained through the measures used in this study.

SPEAKER TYPE	CTTR	CVS1*	SC*	IB	LC
NS (N = 90)	4.00	0.357	2.06	18.86	10.04
	(0.32)	(0.15)	(0.38)	(2.98)	(5.76)
NNS (N = 90)	3.91	0.237	2.31	18.64	11.10
	(0.34)	(0.14)	(0.45)	(3.46)	(5.03)

Table 29: Descriptive statistics in Speaker Type for each measure

5.2 Influence of language teaching experience

While NSs' input has been explored in previous studies, research on the effect of language teaching experience on input modification has been limited. Most research exploring NSs' input modification involved only speakers with language teaching experience (Gaies, 1977; Lynch, 1987), which led to the question of whether speakers with no language teaching experience would exhibit similar input modification when speaking to listeners of varying proficiency levels. The second research question (RQ2) addressed the issue of comparing speakers *with* language teaching experience (WLT) and speakers with *no* language teaching experience (NLT) using the four linguistic and content measures.

5.2.1 Lexical diversity

The first hypothesis under RQ2 (H2a) stated, *speakers WLT will use more lexically diverse speech in the three listener conditions than speakers with NLT. This difference will be indicated by a higher CTTR value for speakers WLT.* The main model indicated a significant difference in CTTR between speakers WLT and NLT. Speakers WLT used more diverse lexis in their speech to all listener type conditions (*M: 3.90 - SD: 0.37*) than speakers NLT (*M: 4.03 - SD: 0.34*). Nevertheless, both speakers WLT and NLT tended to modify their input by using less diverse lexis when speaking to low-proficiency listeners (Figure 28).
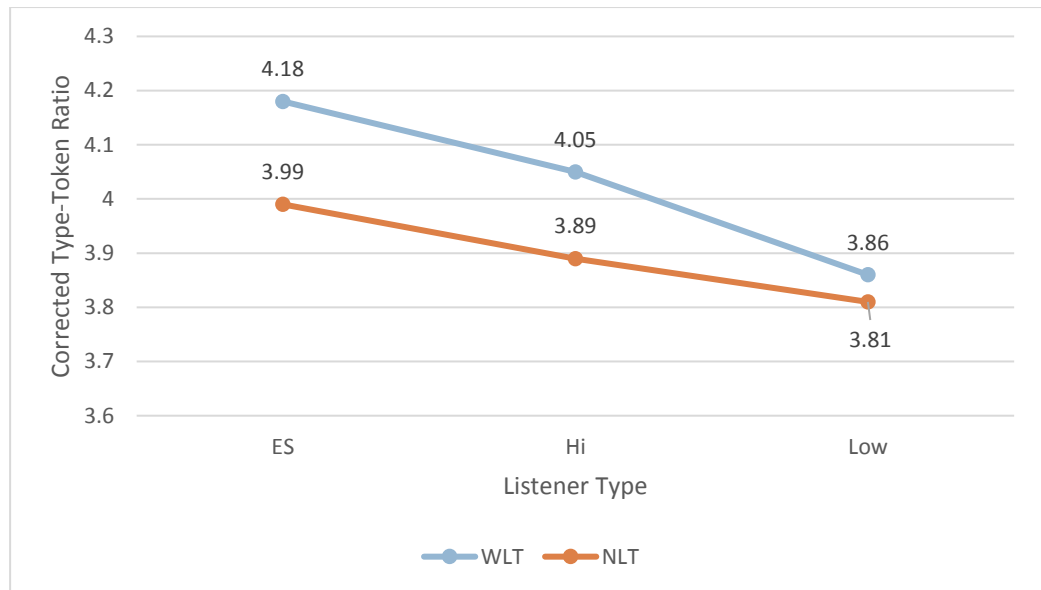
Figure 28: CTTR in Language Teaching condition

It is possible that speakers WLT are more likely to have larger, more diverse lexicons and so be better able to appropriately incorporate more varied lexical items in their input. The speakers with NLT experience, on the other hand, may generally have smaller lexicons or be less inclined to draw on varied vocabulary items in their speech, due to their inexperience in language teaching. In other words, experienced language teachers may have dealt with students of varying proficiency and, therefore, been able to think of different words and phrases to assist students' understandings. They may also be more capable of providing insightful feedback and clarifying answers in response to students' verbal and non-verbal cues. For example, as part of *The Blind Man* (Story B), speakers WLT experience tended to use different vocabulary items referring to "money" in their speech to native listeners: "pocket money," "coins," "change," and "allowance." On the other hand, speakers with NLT experience repeatedly used similar words, such as repeating the word "money" or "coins" throughout their narratives. The following are two excerpts from the transcripts of two native speakers – one WLT and one NLT – both speaking to low-proficiency listeners, in the same lines of Story A, *The Elephant Weight*.

Excerpt E: CTTR in LT (WLT)

NS3 (WLT) speech to low-proficiency listener (Story A)

> **S:** there are men- . there are four men .. and they- ... they want to (a) <u>weigh</u> an
>
> elephant ... they want to know (a) <u>how much the elephant weighs</u> . how (a)
>
> <u>many . kilograms</u> the .. elephant weighs .. ... so they try to put (b) the elephant
>
> on a scale .. (b) <u>he </u>is too big .. (b) <u>the elephant</u> is too big . for the scale . so ..
>
> there is a boy . who .. says- (c) <u>he talks</u> .. and he .. (c) <u> says</u>- he talks ... let's
>
> put the elephant . yes . put the elephant . on a boat-
>
> [lines: 7 – 15]

Excerpt F: CTTR in LT (NLT)

NS4 (NLT) speech to low-proficiency listener (Story A)

> **S:** they want to know how much the elephant (a) <u>weighs</u> ... however . they do
>
> not have a scale . that is (b) <u>big</u> enough to (a) <u>weigh </u>the elephant ... the biggest
>
> scale that they have is not (b) <u>big</u> enough to (a) <u>weigh</u> the elephant . and so
>
> they do not know how they are going to (a) <u>weigh </u>. the elephant ... ... then a
>
> young boy has an idea . and suggests . that they put the elephant into a boat .
>
> [lines: 3 – 11]

In both excerpts, the related vocabulary items are underlined to compare the

speakers' word choice. Even though both speakers were narrating to low-proficiency

listeners, they exemplified different lexical strategies. The speaker WLT used

different forms of "(a) weigh," such as "how much it weighs" and "how many

kilograms," and a different form of the word "(c) says," such as "talks." Also, she

clearly indicated what she meant by "(b) he" by referring to either "the elephant" or

"the boy." The speaker NLT did not exhibit the same lexical diversity; rather, she

repeatedly used the same words, such as "(a) weigh" and "(b) big." These excerpts are

representative of the difference in lexical diversity between speakers WLT and NLT showcased in transcripts from this study.

Speakers WLT and NLT also differed in the feedback they provided to listeners. Although listeners were encouraged to ask questions and request clarifications when needed, many listeners chose not to, which could indicate either a full understanding of the story or simply personal preference. Likewise, all speakers were instructed to provide feedback in response to listeners' requests. However, speakers WLT exhibited noticeably more confirmation checks than speakers NLT, which encouraged listeners to respond and request clarifications. Two different excerpts are presented below to illustrate the type of lexical diversity employed by NNSs WLT and NLT. These excerpts include the listeners' responses, as well, to exemplify the type of interaction between the speaker and the listener and the extent to which feedback was provided, leading to more diverse lexical input. Both excerpts were from different lines of *The Elephant Weight* (Story A)[35] and were narrated to low-proficiency listeners.

Excerpt G: CTTR in LT (WLT)

NNS2 (WLT) speech to low-proficiency listener (Story A)

	Speaker	**Listener**
14	Some group of man	
15		Okay
16	were discussing how to weigh an elephant . you	
17	know the (a) <u>elephant</u> . the (a) <u>animal </u>they want to	
18	weigh it . but the scale is small ... *Do you*	
19	*know the scale?(confirmation check)*	
20		You mean … the
21		uniform?
22	No , the scale is- . I mean the (b) <u>men they are</u>	
23	<u>wearing uniform</u> . because they are (b) <u>officials</u> like	
24	(b) <u>policeman</u> and stuff so they want-	
25		

---

[35] Each narration was different in terms of giving and receiving feedback, and there were not many examples of speakers giving feedback on the same line of the story.

26		What is the meaning
27	(c) <u>Scale</u>.. the (c) <u>machine</u> that you used to (c) <u>weigh</u>	of scale?
28	<u>yourself</u> . to weigh yourself . every month you want	
29	to see am I .. (c) <u>50 kilo</u> .. 57 . kilo 20 . how much	
30	weight I do need to lose or to gain	
31		
32	So that is a scale .. for weighting …	Okay okay
33		
34	So.. they want to weigh an elephant .. okay . and the	Okay
35	scale is small	
36		
37	so that is a problem . they wanted to solve . so ...	Okay
38	they did not know how to solve this problem ..	
39	though . a small kid . a boy came to them and he had	
40	the (d) <u>idea</u> . (d) <u>how to solve this problem</u>	
41		
42	So... his (d) <u>solution</u> or his (d) <u>idea </u>was to first put	Okay
43	the elephant on the boat ...	
44		

Excerpt H: CTTR in LT (NLT)

NNS10 (NLT) speech to low-proficiency listener (Story A)

	Speaker	**Listener**
38	... when they took the elephant out of the boat ..	
39	they filled the boat again- another time with	
40	stones–	
41		*stones?*
42	(a) <u>stone </u>. stone like (a) <u>hard object</u> like the one	(listener interrupted)
43	you find in the ground in the soil .. in the- if you	
44	are walking by and you ... *okay just a minute let*	
45	*me think about (struggles to explain)*	
46		A sound of water like ??
47	the stone is like a .. it is a hard object . *hard* . *you*	
48	*know* . (a) it is <u>not soft</u>-	
49		…
50	stones	
51		stone ah okay .. like the
52		small thing of-
53	yes. exactly . it is the (a) <u>stones</u> . because it is	
54	. small ... so they have stones . they fill the boat	
55	with stones ... and the boy went down on the boat	
56	. and they kept- the government people . kept	
57	filling the boat . with (a) <u>stones</u> until .. it is the	
58	same weight as the elephant ... the boy knew that	
59	. because the water came to the same mark he put	

60	. . okay? ... then when- when the ama- when the	
61	(a) <u>stones</u> in the boat is the same weight as the	
62	elephant .. they started taking (a) <u>the stones</u> . and	
63	put them on the scale .. and by that they knew the	
64	weight of the elephant ... that is it	
65		elephant is a type of the
66		fish?
67	no the elephant *is you know-*	
68		Elephant has .. (gestured
69		with hand)
70	yes exactly . it is the one that has a long nose	
71		Okay I got it

These excerpts show the extent to which NNSs WLT and NLT attended to listeners' cues and that the NNS WLT was better able to respond to listener feedback effectively. For example, in excerpt G, the NNS WLT narrated part of the story, followed by a confirmation check: "*Do you know the scale?*" (lines 18-19). The listener clearly did not understand what "scale" meant, so the speaker responded accordingly. She was also able to explain and clarify the meanings of certain words by using synonyms and paraphrases. She referred to (a) "elephant" as "animal," (b) "men in uniforms" as "officials" and "policeman," (c) "scale" as "machine to weigh yourself," and (d) "idea" as "solution." Interestingly, the use of different words to support listener comprehension may explain in part the greater lexical diversity in the speech of speakers WLT.

On the other hand, the NLT speaker did not exhibit the same level of attention to the listener's cues. As excerpt H illustrates, the listener interrupted the speaker to clarify the meaning of the word "stones" (lines 40-41). The speaker responded by referring to (a) "stones" as "hard objects" and struggled (lines 44-45) and eventually failed to give different forms of the word or phrases so the listener could understand. The speaker then continued to use the word "stones" throughout the narrative, indicating the speaker's inability to use diverse lexis to support listener

comprehension. Finally, even though the whole story revolved around the elephant, by the end of the story, the listener (lines 65 – 66) was still not sure what was meant by "an elephant." In turn, the speaker responded by saying, "the elephant .. you know-" and the listener had to gesture that he understood what it was (lines 67 – 70).

All of the above excerpts showed that speakers WLT provided more feedback to listeners, using different words or phrases to support listener comprehension and drawing on their experience in language teaching to influence the way they narrate the story. This was true for both NSs and NNSs. Both NSs and NNSs NLT, however, did not attend to listener cues as skillfully and when clarifying word meanings, often failed to explain the meaning using different words. These trends in readiness to provide feedback may have contributed to the difference in lexical diversity between speakers WLT and NLT.

5.2.2 Lexical sophistication
The second hypothesis under RQ2 (H2b) stated, *speakers WLT will use more lexically sophisticated speech in the three listener conditions than speakers with NLT. This difference will be indicated by a lower CVS1 value for speakers WLT*. The main model found no significant difference between speakers WLT and NLT in CVS1. That is, speakers WLT (*M: 0.310 - SD: 0.15*) did not differ from speakers NLT (*M: 0.285 - SD: 0.16*) in lexical sophistication. Descriptively, however, speakers WLT did use more sophisticated words than NLT with all listener types, indicating a subtle difference between WLT and NLT in the use of low-frequency words. As exhibited in Figure 29, even though WLT means were higher, both WLT and NLT tended to use less sophisticated words with non-native listeners than with native listeners.
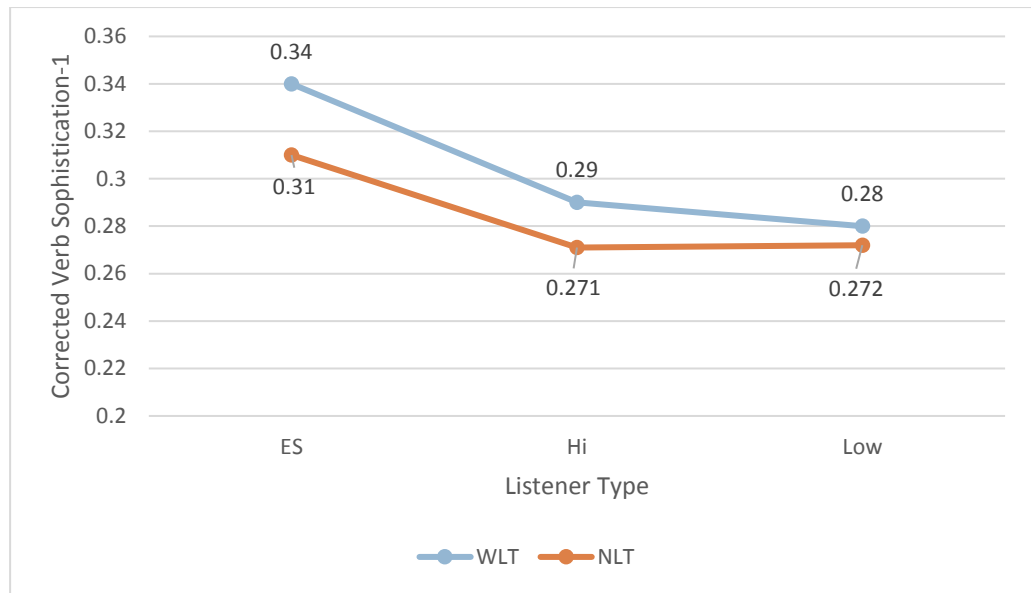
Figure 29: CVS1 in Language Teaching condition

5.2.3 Syntactic complexity

The third hypothesis under RQ2 (H2c) stated, *speakers WLT will exhibit more complex syntax in their speech in the three listener conditions than speakers with NLT. This difference will be indicated by a lower number of S-nodes per clausal utterance for speakers WLT*. The model found no significant difference between speakers WLT (*M: 2.24 - SD: 0.44*) and speakers NLT (*M: 2.13 - SD: 0.42*) in SC. Descriptively, WLT seemed to use more complex utterances than NLT with all listener groups, indicating a non-significant difference between WLT and NLT in their use of syntax. Both WLT and NLT tended to employ shorter, less complex utterances with non-native listeners than with native listeners (Figure 30).
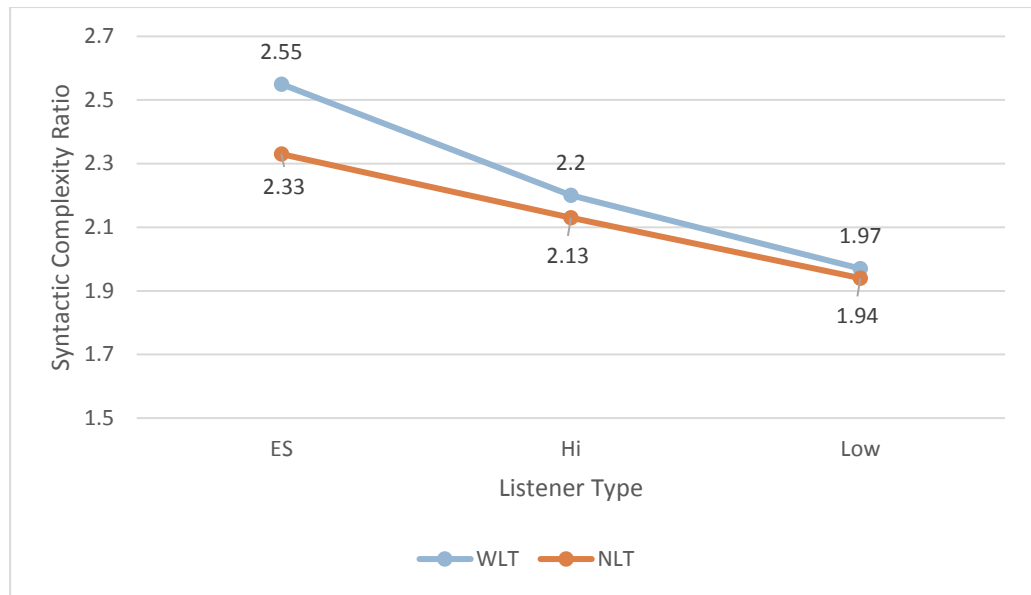
Figure 30: SC in Language Teaching condition

5.2.4 Content dilution

Finally, the last hypothesis under RQ2 (H2d) stated, *speakers WLT will include more content details in their speech in the three listener conditions than speakers with NLT. This difference will be indicated by higher counts of mentioned IBs for speakers WLT*. The model found no significant difference between speakers WLT (*M: 19.20 - SD: 2.95*) and speakers NLT (*M: 18.31 - SD: 3.43*) in IBs. Descriptively, WLT included more information than NLT with all listener types, indicating a non-significant difference between WLT and NLT in the content mentioned. In addition, both WLT and NLT tended to omit more IBs when speaking to non-native listeners than to native listeners (Figure 31).
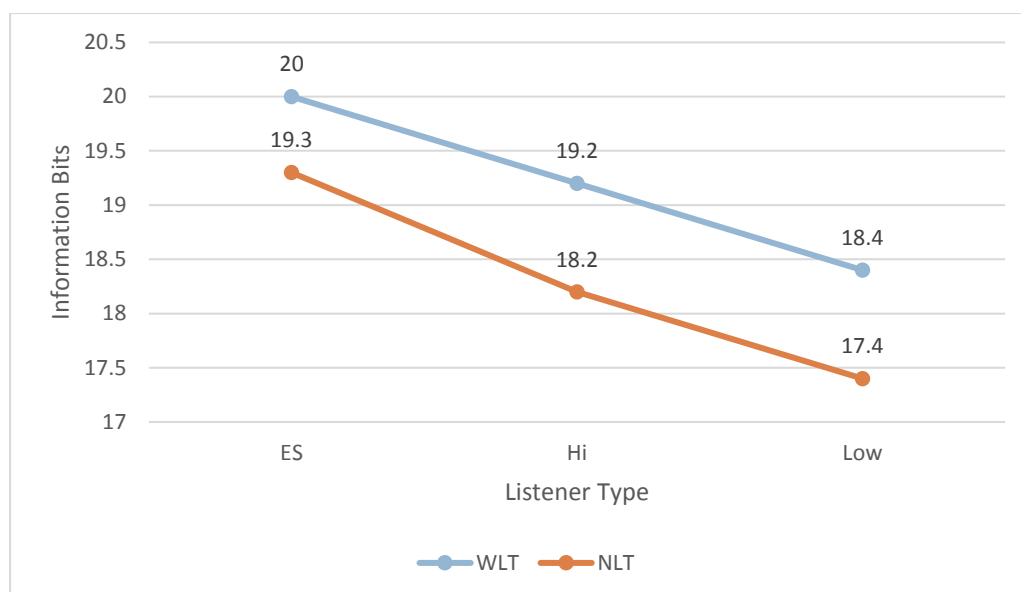
Figure 31: IB in Language Teaching condition

5.2.5 Summary

Speakers WLT and speakers NLT differed significantly on only one measure – lexical diversity – and speakers WLT demonstrated greater lexical diversity than speakers NLT. Though descriptively different, measures for lexical sophistication, syntactic complexity, and content dilution showed no statistically significant differences between speakers WLT and NLT. Language teaching also showed no effect on listener content comprehension. Table 30 summarizes these results.

LANGUAGE TEACHING	CTTR*	CVS1	SC	IB	LC
WLT (N = 90)	4.03	0.310	2.24	19.20	10.82
	(0.34)	*(0.15)*	*(0.44)*	*(2.95)*	*(5.47)*
NLT (N = 90)	3.90	0.284	2.13	18.31	10.32
	(0.37)	*(0.16)*	*(0.42)*	*(3.43)*	*(5.38)*

Table 30: Descriptive statistics in Language Teaching for each measure

5.3 NSs' speech towards non-native listeners

This section discusses the differences found in NSs' speech towards the three listener conditions. The discussion addresses the third research question (RQ3), which related to whether NSs, regardless of language teaching experience, modify their

speech in terms of the four linguistic and content measures, according to listeners'

proficiency levels.

5.3.1 Lexical diversity

The first hypothesis under RQ3 (H3a) stated, *NSs will use less lexically*

diverse speech with low-proficiency listeners. This difference will be indicated by a

lower CTTR value for low-proficiency listeners. The NS-only model found a

significant difference for CTTR between ES and Hi and between ES and Low. This

result indicated that the NSs used more diverse lexis when speaking to native controls

(*M: 4.13 - SD: 0.31*) than to high-proficiency listeners (*M: 4.02 - SD: 0.27*) or low-

proficiency listeners (*M: 3.83 - SD: 0.32*). This is understandable, as the listeners in

the high-proficiency sample were not truly advanced but, rather, intermediate, as

explained in Chapter 3. The following two excerpts illustrate the type of differences in

CTTR between the same NS's speech to an ES and to a low-proficiency listener. They

are both from the same lines of Story B, *The Blind Man*.

Excerpt I: CTTR in NS speech (ES)

NS5 (NLT): speech to ES listener (Story B)

> **S:**…so there is a little boy who has .. some money . and (a) <u>he really wants to</u>
>
> <u>go buy a toy for himself</u> . and he goes to the toy store ... but when he is at the
>
> toy store . (b) <u>he sees an old man .. who's blind . begging for money on the</u>
>
> <u>street</u> ... and the little (c) <u>boy starts to feel bad</u> for the little- for the old man .
>
> and (d) <u>he decides that instead of spending the money on a toy for himself</u> . <u>he</u>
>
> <u>is going to give his money to the old man</u> … so he crosses the street .. to (e)
>
> <u>put</u> the money in the old man's tin . but as he (e) <u>drops</u> the money into the tin .
>
> a woman gets out of a car . and slams the door shut so it makes a (f) <u>loud bang</u>
>
> ... so the old man thinks that it was the woman .. who dropped the money . in
>
> his tin because he turned where he heard the (f) <u>noise</u> .

114

Excerpt J: CTTR in NS speech (Low)

NS5 (NLT): speech to low-proficiency listener (Story B)

> **S:** There is a little boy … he has some money ... and (a) he wants to buy a toy .
>
> but when he goes to the toy store . (b) he sees an old man .. who's blind .. the
>
> old man cannot see . and (c) the little boy feels bad .. (c) he feels really upset
>
> .. for this old man … so . (d) instead of buying himself a toy . he decides to (f)
>
> give his money to the blind man ... but whenever he (e) drops his money .. into
>
> the old man's can .. a woman . gets out of a car . and slams the door … so the
>
> old man .. thinks.. that the woman . (f) gave him the money . and not the little
>
> boy

The related phrases in both excerpts were underlined to compare the NS's

word choice when speaking to a native listener and to a low-proficiency listener. For

instance, with the ES, she used (e) "puts" and "drops" and, instead of repeating the

phrase (f) "loud bang," she used the word "noise." On the other hand, with the low-

proficiency listener, she repeated the word (c) "feels" twice – "the boy feels sad" and

"feels really upset" – as well as the word (f) "give" – "he decided to give" and "the

woman gave him." These qualitative examples are reflected in the CTTR values: a

ratio of 4.01 for the ES and a ratio of 3.37 for the low-proficiency listener.

In addition, examining the narrations by the same speaker of the same story to

different listener types provided a different level of analysis of lexical diversity.

Lexical diversity does not mean only different words with the same meaning; it can

also denote the elaboration of a segment of the story or content. For instance, when

describing what the boy thinks, the NS said to the native listener that the boy (d)

"decides that instead of *spending the money on a toy for himself*. he is going to give his money to the old man," compared to her statement to the low-proficiency listener that (d) "instead of *buying himself a toy* . he decides to give his money to the blind man." The former utterance incorporated more terms in describing the same content, while the latter was simple and straightforward: the simpler "*buying* himself a toy" rather than the more complex "*spending the money* on a toy for himself." The latter indicates a higher level of lexical sophistication.

The speaker also told the ES that the boy (a) "really wants to go buy a toy for himself" and (b) "he sees an old man .. who's blind . begging for money on the street," but told the non-native listener that (a) "he wants to buy a toy" and (b) "he sees an old man .. who's blind .. the old man cannot see." While she clarified what she meant by *blind* in her speech to the low-proficiency listener, she described what the blind man was doing, *begging*, in her speech to ES. This example suggests that in modifying her input to a low-proficiency listener, by clarifying word meaning, led to loss of content information, as the speaker emphasized the meaning of *blind* and discarded the key detail of *asking for money*.

These findings are consistent with earlier research wherein NSs generally modified their input by increasing word repetition when teaching or speaking to intermediate or low-proficiency learners (Chaudron, 1983a, 1983b; Schierloh & Paulsell, 2010). Earlier findings demonstrated that the repetition of simple words helped L2 learners remember those words (Chaudron, 1983a, 1983b), and so it had been seen as a supportive feature for language learning. In this study, however, repetition and reduced lexical verity led to content dilution and did not assist listeners in learning the content of the stories. This is discussed further in section 5.5.

5.3.2 Lexical sophistication

The second hypothesis under RQ3 (H3b) stated, *NSs will use less lexically sophisticated speech with low-proficiency learners. This difference will be indicated by a lower CVS1 value for low-proficiency listeners.* The NS-only model (using the transformed data) found a significant difference in CVS1 between ES and Low, but no difference between ES and Hi. This result indicated that the NSs used less sophisticated lexis with the low-proficiency listeners (*M: 0.25 - SD: 0.13*) than with the native controls (*M: 0.33 - SD: 0.14*). High-proficiency listeners, interestingly, received slightly more sophisticated speech (*M: 0.34 - SD: 0.11*) than ES, although the difference was not statistically significant. The following two excerpts illustrate the types of differences in CVS1 between the same NS's speech to an ES and to a low-proficiency listener. They are both from the same lines of Story C, *The Indian Man*.

Excerpt K: CVS1 in NS speech to listeners (ES)

NS7 (NLT): speech to ES listener (Story C)

S: he went to look for help .. a small boy (a) came up to him . and began to talk ... while they were talking a man with dark glasses . (b) went behind him and (c) grabbed the basket ... the man with the dark glasses then ran away with the basket . and the boy ... the Indian men (d) saw a police officer nearby . with the whistle . who tried to stop the thieves . but it was (e) unsuccessful ... the man and the boy (f) hopped into a car and drove away

[lines 4-11]

Excerpt L: CVS1 in NS speech to listeners (low)

NS7 (NLT): speech to low-proficiency listener (Story C)

S: so he went to look for help ... when he turned around a boy (a) approached him . and they began to talk ... a man with dark glasses . (b) came behind the

117

Indian . and (c) <u>took</u> his basket ... the thief ran away with the basket . and the boy joined him ... the Indian (d) <u>alerted</u> . a policeman who blew a whistle . but (e) could not make the thieves stop ... the thieves (f) <u>got</u> into the car . and drove away .

[lines 3 - 10]

The related phrases in both excerpts were underlined to compare the NS's lexical sophistication when speaking to an ES and to a low-proficiency listener. With ES, for instance, she used sophisticated words such as (c) "grabbed," (e) "unsuccessful," and (f) "hopped," while with the low-proficiency listener, she used words such as (c) "took," (e) "could not make the thieves stop," and (f) "got." This suggests that the speaker's sensitivity to the listener's proficiency prompted her to modify her language by using more common words. In addition, the speaker used the phrasal verb (a) "came up to" with the native listener but chose to use the word (a) "approached" with the non-native listener. The phrasal verb "came up to" is, arguably, considered advanced and is usually used only among NSs and advanced NNSs. Earlier findings indicated that speakers use more collocations or idiomatic expressions in their speech to NSs compared to NNSs (Chaudron, 1982; Henzl, 1973, 1979; Kliefgen, 1985; Lynch, 1987; Mizon, 1981). Other examples of such expressions used by NSs in this study include "step off," "made a noise," "got upset" "act of kindness," and "pangs of conscience."

Interestingly, the speaker used (d) "alerted" with the low-proficiency listener and (d) "saw" with the native speaker. This indicates that although speakers may use sophisticated words with non-native listeners, it tends to be less frequent than with native listeners. These qualitative examples were reflected in each transcript's CVS1 value: a ratio of 0.42 for ES and a ratio of 0.29 for Low.

These results are consistent with earlier findings in the SLA literature. Similar to lexical diversity, research on NSs' speech to NNSs has generally found lower lexical sophistication (Chaudron, 1988; Gaies, 1977; Lynch, 1987). In language classroom settings, instructors tended to use high-frequency words to assist students' comprehension (Krashen, 1981, 1982). However, exchanging low-frequency words for high-frequency words may not work efficiently in content classrooms (Schierloh & Paulsell, 2010). Language in CLIL and EMI environments is not flexible and as such requires the instructor to use certain vocabulary and phrases in specific contexts. If specific vocabulary items are replaced by simpler, more frequent words, the quality of the subject matter instruction will likely be affected.

5.3.3 Syntactic complexity

The third hypothesis under RQ3 (H3c) stated, *NSs will exhibit less complex syntax in their speech with low-proficiency listeners. This difference will be indicated by a lower number of S-nodes per clausal utterance for low-proficiency listeners.* The NS-only model found a significant difference in NSs' speech between ES and Low and between ES and Hi. These results indicate that NSs used more complex syntax when narrating the story to native listeners (*M: 2.58 – SD: 0.44*) than to either high-proficiency listeners (*M: 2.29 – SD: 0.38*) or low-proficiency listeners (*M: 2.05 – SD: 0.36*). The following two excerpts from the same NS's speech to an ES and a low-proficiency listener elucidate the type of differences found in SC between the two listener groups. Both come from the same lines of Story A, *The Elephant Weight*.

Excerpt M: SC in NS speech to listeners (ES)

NS1 (WLT): speech to ES listener (Story A)

> **S:** /next story :: takes place at a . shipyard or kind of like a dock/ (1/1) .. /and there :: are some government officials *who* :: want to . :: weigh an elephant to :: see .. *I guess* how much it :: weighs for whatever reason/ (5/1).. /and they are

:: using a scale to :: weigh it . but then they :: realize the elephant :: is too big .

to :: put on the scale/ (5/1).. /and .. so they :: are kind of stuck there/ (1/1) ..

/but then a boy :: thinks . of an alternative way to :: weigh the elephant/ (2/1)

    [lines 1-8]

Excerpt N: SC in NS speech to listeners (low)

NS1 (WLT): speech to low-proficiency listener (Story A)

S: /in this story . there :: is an elephant/ (1/1).. /and the .. and there :: are also

some government officials/ (1/1).. /the . government officials :: want to ::

know the weight of the elephant/ (2/1). but because the elephant :: is . very

heavy they :: want to . :: weigh it/ (3/1) .. /so they . they :: put the elephant on

a scale . but the elephant :: is too big/ (2/1) .. /so they cannot :: find the weight

of the elephant/ (1/1).. / the scale :: is too small/ (1/1).. /so then a boy :: has an

idea . about a better way to :: weigh the elephant/ (2/1)

    [lines 1-9]

While excerpt M, to an ES, included a total of fourteen S-nodes and five

clausal utterances (SC ratio: 2.8), excerpt N, to a low-proficiency listener, included a

total of 13 S-nodes and eight clausal utterances (SC ratio: 1.62). While the NS's

speech to ES used more syntactically complex utterances overall, his speech to the

low-proficiency listener used less complex syntax overall. He mostly incorporated

shorter, simpler clauses (e.g., in excerpt N, five out of eight utterances were simple

ones) and tended to use coordinating conjunctions more often than subordinate ones.

This example illustrates the extent to which one speaker modified his input by using

simple, short utterances with the low-proficiency listener and using longer, more

complex utterances with the native listener. On another note, instead of using a

pronoun, the speaker repeated "the elephant" seven times to the non-native listener to

clarify his reference in most of the mentioned details. For his speech to ES, however, he only mentioned "the elephant" three times and instead used "it" in reference to *the elephant*, indicating less repetition (high lexical diversity) used with ES compared to the low-proficiency listener.

These results corroborate earlier findings in which NSs generally simplified their input with low-proficiency learners (Chaudron, 1988; Ferguson, 1971; Gaies, 1977; Wesche & Ready, 1985; Schierloh & Paulsell, 2010). Gaies (1977) found that instructors' syntax became more complex when geared towards advanced learners and less complex when geared towards low-proficiency learners. In a more recent study, Schierloh and Paulsell (2010) found that teachers in a business course used simplified syntax and increased repetitions, similar to what was noted above regarding the repetition of "the elephant." The NSs' simplification of their syntax reflected their perceptions of the listeners' proficiency and represented an attempt to assist the listeners in understanding the main concepts of the stories. However, such simplification of syntax does not necessarily work well in content learning. The learners' understanding of the content is primarily contingent upon their proficiency level. Simplification of syntax and lexis leads to another type of input modification: content dilution.

5.3.4 Content dilution
Finally, the last hypothesis under RQ3 (H3d) stated, *NSs will omit more content details in their speech with low-proficiency listeners. This difference will be indicated by lower counts of IBs mentioned for low-proficiency listeners*. The NS-only model found significant differences in IBs between ES and Low and between ES and Hi. This indicates that native speakers tended to omit more information when speaking to high- (*M: 18.96 - SD: 2.70*) and low-proficiency listeners (*M: 17.46 - SD: 2.77*) than when speaking to the native controls (*M: 20.16 - SD: 2.63*).

Additionally, NSs tended to omit the most content information when speaking to low-proficiency listeners. In *The Indian Man*, for instance, a speaker omitted the fact that the policeman "blew his whistle"; instead, the speaker told a low-proficiency listener that the policeman "yelled." In another example, the speaker mentioned a "toy store" to a native listener, but only a "store" to a non-native listener. These examples and others were compared within speakers and across listeners to examine instances of speakers omitting content information. To further demonstrate these types of differences, the following two excerpts from the same NS's speech to an ES and to a low-proficiency listener further demonstrate these types of differences. They are from the same lines of Story B, *The Blind Man*.

Excerpt O: IB in NS speech to listeners (ES) (I)

NS9 (NLT): speech to ES listener (Story B)

> **S:** (1) :: he decides to go over to the blind man who is begging for change ... and . (2) :: as he is walking over . somebody .. pulls up next to the blind man . in a car ... so . (3) ::when the woman gets out of the car . (4) :: she . shuts the door .. and (5) ::at the same time .. (6) ::the boy puts change (7) ::into the blind man's jar ... but the blind man since (8)::he heard the door slam . (9)::thinks (10) ::that someone got out of the car .. and put money in his change jar . but- instead . it was the boy .. and the boy is sad .. that- and (11)::the man like gestures with his hat .. (12)::to say . thank you ... and then . so- but (13)::the boy is disappointed (14)::because . he- the blind does not know . that he is the one . that put the change in the jar.
>
> > [lines 9-22]

Excerpt P: IB in NS speech to listeners (low) (I)

NS9 (NLT): Speech to low-proficiency listener (Story B)

S: (1) ::he decides . to go over . and give . the blind man his change ..

(2)::when the boy is giving the man his money ... a car stops and

(3)::somebody gets out of the car .. so (4)::the blind man .. hears the car stop .

and the door of the car . close . and (5)::thinks . that (6)::the person who has- .

who has stopped in the car . was the one who is giving him the change .. and

(7)::says thank you ... but . (8)::the boy is sad because . he is actually the one

that gave the blind man the money.

[lines 4-17]

Both excerpts were segmented by double colons to mark the IBs and were numbered for the purpose of describing the missing information. While the excerpt to an ES included 13 IBs (story total: 23), the excerpt to a low-proficiency listener included only eight IBs (story total: 17). The speaker altered his speech to the low-proficiency listener by omitting IBs such as *the woman shuts the door, the boy puts the change, he puts it in the blind man's jar*, and *the man gestures with his hat* (propositions 4 to 7 and 11 in Excerpt O). In addition, the speaker did not mention that the boy was upset because "the man didn't recognize it was him" and "his good deed was not recognized." Instead, he stated that the boy was upset "because he is the one that gave the blind man the money," a vague statement that could lead to a different interpretation (e.g., the boy regrets giving the money). However, these IBs were mentioned in his narration to the native listener, suggesting that the content was diluted by his input modification when speaking to the non-native listener. Since content dilution is a relatively newly operationalized concept in the SLA literature, two more excerpts are given below, using a different story with another NS, in order to further illustrate its role in this study.

Excerpt Q: IB in NS speech to listeners (ES) (II)

NS3 (WLT): speech to ES listener (Story C)

> **S:** (1)::he has a briefcase and a large square basket .. he- I think (2)::the basket
>
> is really heavy so (3)::he is looking for someone to help him carry it . but
>
> while he is looking . (4)::a boy comes up to him and (5)::starts talking to him
>
> ... and while he is talking to the boy (6)::a man with a hat and sunglasses ..
>
> (7):: picks up the basket while the ma - while the Indian man's not looking .
>
> and (8)::runs off with it . and (9)::the boy goes with him ... (10)::the Indian
>
> man gets the attention of a cop and ask- and (11)::the policeman blows his
>
> whistle . and tries to get the .. man with the sunglasses and the boy . who have
>
> the basket to stop .
>
> [lines 2 - 13]

Excerpt R: IB in NS speech to listeners (Low) (II)

NS3 (WLT): speech to low-proficiency listener (Story C)

> **S:** the Indian man (1) :: has .. a large basket .. a large . square . basket ... he- .
>
> (2) ::a boy .. starts talking . to the Indian man
>
> **L:** Start talking?
>
> **S:** He talks to- the boy talks to the Indian man .
>
> **L:** (…)[36]
>
> **S:** So the boy talks .. to the Indian man ... while the boy and the Indian man
>
> are talking .. (3)::another man . with a hat.. and sunglasses .. (4):: takes the
>
> basket .. steals the basket ... and (5)::that man . with the hat and sunglasses ..
>
> that man . and the boy .. (6)::run away .. with the basket . they steal the basket
>
> together ... so the man and the boy steal the basket from the Indian man ... the

---

[36] (…) indicates a response to a question raised by the listener. These segments were excluded to keep the excerpt short. These excluded segments did not have new content to be counted as IB; rather, the speaker repeated and/or further clarified parts of the story.

Indian man- …. (7)::the Indian man .. tells the police (…). the Indian man asks

the police for help ... and (8)::the police .. blow a whistle . and try to stop- the

police try to stop (9)::the thieves . the man who took the basket ...

[lines 5-29]

While the excerpt to an ES included 11 IBs (story total: 21), the excerpt to a

low-proficiency listener included only nine (story total: 16). In narrating the story to

the low-proficiency listener, the speaker modified her speech by omitting IBs such as

the basket was heavy, the Indian man was looking for someone to help him, and *the*

boy came to the Indian man (propositions 2 to 4 in Excerpt Q). Including the fact that

t*he boy came up to the Indian man before talking to him* gives a clearer understanding

of how the story unfolds; excluding it, however, introduces information gaps that are

often filled with false interpretations. This does not mean that information was only

mentioned to ESs and omitted for non-native listeners. In some cases, speakers

mentioned details to non-native listeners and excluded them in their speech to ES. For

instance, in the above excerpts, the speaker mentioned "thieves" (example 9 in

excerpt R) only to the low-proficiency listener, indicating a distinct alteration.

However, the general results indicated that native listeners received significantly more

content information than non-native listeners. These results demonstrate not only the

occurrence of simplified input but also suggest that NSs diluted content more when

speaking to non-native listeners of both high- and low-proficiency than when

speaking to native listeners.

5.3.5 Summary
Corroborating earlier findings, this study found that NSs modified their input

according to listener type. Linguistic modification was found using three linguistic

measures to gauge the levels of lexical diversity, lexical sophistication, and syntactic

complexity employed in NSs' speech to the listener conditions (Table 31).

125

Importantly, a content measure also indicated a significant content dilution when NSs were speaking to high- and low-proficiency listeners. These listeners received not only simplified input but also content that was so altered that some propositions were omitted. This may cause the non-native listener scores to be lower on the content comprehension measure (discussed in section 5.5).

Linguistic and content measures in NSs' speech				
Listener Type	**CTTR**	**CVS1**	**SC**	**IB**
ES	4.13	0.33	2.58	20.16
HI	4.02*	0.34	2.29*	18.96*
LOW	3.83*	0.25*	2.05*	17.46*

Asterisks represent statistically significant differences in comparison to ES

Table 31: Linguistic and content measures in NS's speech

5.4 Advanced NNSs' speech towards non-native listeners

As discussed earlier, significant differences were found between NSs and NNSs in lexical sophistication and syntactic complexity; they were similar in both lexical diversity and content dilution. Additionally, both speaker types showed a similar pattern in their speech towards native and non-native listeners. Considering that the majority of instructors in EMI settings are NNSs, it is important to describe the differences found within NNSs' speech towards the three listener conditions, as is posed by the fourth research question. RQ4 set out to explore further whether advanced NNSs, regardless of language teaching experience, modify their speech when narrating the stories to low- and high-proficiency listeners, compared with ESs, based on the four linguistic and content measures.

5.4.1 Lexical diversity

The first hypothesis under RQ4 (H4a) stated, *NNSs will use less lexically diverse speech with low-proficiency listeners. This difference will be indicated by a lower CTTR value for low-proficiency listeners.* The NNS-only model found a significant difference in NNSs' speech between ES and low-proficiency listeners.

This indicates that low-proficiency listeners received less diverse lexis from NNSs (*M: 3.83 - SD: 0.38*) than the native controls did (*M: 4.03 - SD: 0.40*). There was, however, no significant difference between ES and high-proficiency listeners (*M: 3.91 - SD: 0.39*). The following two excerpts from the same NNS's speech – one to an ES and one to a low-proficiency listener – illustrate the type of differences found in CTTR according to listener type. Both are from the same lines of Story C, *The Indian Man.*

Excerpt S: CTTR in NNS speech (ES)

NNS10 (NLT): speech to ES (Story C)

> **S:** the Indian man noticed he called for the (a) policeman . who has a whistle
>
> ... he whistles so that people can catch (b) the thieves . the thieves (c) got away
>
> . he could not . catch them ... (d) they got into a car . and drove . next to a zoo
>
> .. into a woods- after the zoo there is wood there is a forest . they went inside .
>
> they stopped the car because they felt safe that nobody can catch them at that
>
> point .. they opened the bag . to their surprise there came a big snake
>
> > [lines 14 - 23]

Excerpt T: CTTR in NNS speech (Low)

NNS10 (NLT): speech to low-proficiency listener (Story C)

> **S:** the Indian .. tried to- he noticed by them .. so he tried to reach for the (a)
>
> policeman .. he saw the (b) policeman
>
> **L:** what?
>
> **S:** he tried to call the (c) policeman ... so that he can catch the thieves ... so
>
> when he .. told the (d) policeman . the (e) policeman has a whistle ... he
>
> whistled . so that people can get attention and they catch the thieves . but they
>
> could not catch them ... they (f) got away .. they rode in a car .. (g) the man

with the sunglasses and the boy . they rode- they got into the car . and they

dro- drove ... as they are (h) <u>driving away</u> .. now the (i) <u>policeman</u> cannot

catch them they already (j) <u>went away</u> . okay they passed by a zoo ... after the

zoo there is woods there is trees ... you know the woods?

L: yes .

S: . okay so they- . when they reached the woods . they felt that they already .

are safe . no (k) <u>policeman</u> are following them . they stopped the car ... they

went down and opened the bag ... and to their surprise there was a big snake

[lines 40 - 61]

The related words in the excerpts are underlined to demonstrate the lexical

diversity used by the same NNS with a native listener compared to a low-proficiency

listener. Even though both excerpts refer to the same scene in the story, the excerpt to

the low-proficiency learner is much longer and has 75 word types and 164 word

tokens (CTTR: 3.7); the excerpt to ES is shorter and has 53 word types and 86 word

tokens (CTTR: 3.96). The CTTR for these excerpts suggests that length does not

determine lexical diversity, because increased repetition of similar words in the

second excerpt led to a lower TTR, i.e., less diverse speech. In these samples, there

were distinct differences in the amount of detailed repetition directed to the different

listener types. For instance, the speaker only mentioned the (a) "policeman" once in

his narration to the native listener, but mentioned it at least seven times in his speech

to the non-native listener. He also repeated "got away" or "driving away" three times

in his speech to the non-native listener and only once to the native listener.

Interestingly, the word "thieves" was mentioned in both excerpts, but the

speaker clarified to the non-native listener that the *thieves* were (g) "the man with the

sunglasses and the boy," whom he had mentioned a few times earlier in the story.

These examples illustrate the type of lexical input provided by NNSs to the different groups of listeners. NNSs' speech exhibited results similar to the NSs', as both speaker types tended to repeat lexical items – i.e., use less diverse lexis – when speaking to low-proficiency listeners than to native controls.

Nevertheless, unlike the NSs, the NNSs did not demonstrate a significant difference between ES and high-proficiency listeners. This may have occurred because NSs have a wider range of vocabulary and were therefore statistically able to show the difference. Although there has been limited research on input modification by NNSs, the findings here verify that NNSs are capable of modifying their input by using less diverse lexis when speaking to listeners of different proficiency levels – at least differentiating NS and NNS listeners. NNSs in this study increased their word repetition when their input is geared towards low-proficiency learners, suggesting that advanced NNSs display a sensitivity to the listener's proficiency similar to that of NSs.

5.4.2 Lexical sophistication

The second hypothesis under RQ4 (H4b) stated, *NNSs will use less lexically sophisticated speech with low-proficiency learners. This difference will be indicated by a lower CVS1 value for low-proficiency listeners.* The NNS-only model (using the transformed data) showed a significant difference in NNSs' speech between ES and low-proficiency listeners, but no difference between ES and high-proficiency listeners. Low-proficiency listeners received significantly less sophisticated lexis from NNSs (*M: 0.19 - SD: 0.11*) than the native controls did (*M: 0.27 - SD: 0.15*). High-proficiency listeners received slightly less lexically sophisticated speech (*M: 0.22 - SD: 0.13*) than ES listeners, although this difference was not significant.

To illustrate the type of differences found in CVS1, the following two excerpts are reproduced from the same NNS's speech to an ES and to a low-proficiency listener. They are from the same lines of Story B, *The Blind Man*.

Excerpt U: CVS1 in NNS speech to listeners (ES)

NNS7 (WLT): speech to ES (Story B)

> **S:** he crossed the (a) road ... and as he was . putting the money into the tin of the man . giving him the money . a lady stopped her car . nearby . and she got off the car and she (b) slammed the door ... so the blind man . (d) associated the (e) noise that he heard slamming of the door . with the money that he got ... he thought that the person who got off the car is the one who gave him the money ... so what he did is he . took off his hat . in appreciation as- as . gesturing- as a (c) gesture for appreciation to the person who got off the car . of giving him the money .
>
> > [lines 14 - 24]

Excerpt V: CVS1 in NNS speech to listeners (low)

NNS7 (WLT): Speech to low-proficiency listener (Story B)

> **S:** the boy . crossed the (a) street and as he was giving hi- . the man mone- the money and putting the money in his . tin .. a woman . stopped . her car and she got off the car then she (b) closed the door .. the beggar- the man … he thought the woman who got off the car is the one who gave the money for him ... so he . removed his hat and as a (c) sign of . appreciation . for the woman ..
>
> > [lines 9 - 16]

The related words in the excerpts were underlined to demonstrate the NNS's word choice in his speech towards the ES compared to the low-proficiency listener. With the ES listener, for instance, he used words such as (a) "road," (b) "slammed,"

and (c) "gesture," compared to words such as (a) "street," (b) "closed," and (c) "sign" with the low-proficiency listener. This adjustment indicates the NNS's sensitivity to listener proficiency, which triggered him to modify his input by incorporating less sophisticated words, i.e., more high-frequency lexis. In addition, the speaker included words such as (d) "associated" and (e) "noise" when speaking to the native listener and chose to omit these words in his speech to the non-native listener. Here, the NNS simplified his input to get the main idea across, providing less sophisticated input and diluted content by excluding words and IBs. Besides their ability to use less *diverse* lexis with low-proficiency listeners, as discussed in the previous section, these results prove that advanced NNSs can also modify their speech by using less *sophisticated* lexis. These qualitative examples reflect the type of lexical input provided by NNSs, which exhibited results similar to the NSs. Although NSs and NNSs differed in their levels of lexical sophistication, both modified their input by using less sophisticated words with non-native listeners, especially those with low proficiency.

5.4.3 Syntactic complexity

The third hypothesis under RQ4 (H4c) stated, *NNSs will exhibit less complex syntax in their speech with low-proficiency listeners. This difference will be indicated by a lower number of S-nodes per clausal utterance for low-proficiency listeners.* The NNS-only model found a significant difference in NNSs' speech between ES and low-proficiency listeners, as well as between ES and high-proficiency listeners. The NNSs used more complex syntax with native listeners (*M: 2.30 – SD: 0.37*) than with high-proficiency listeners (*M: 2.04 – SD: 0.35*) or low-proficiency listeners (*M: 1.85 – SD: 0.28*). To clarify the type of differences in SC between ES and low-proficiency, the following two excerpts are reproduced from the same NNS's speech to an ES and to a low-proficiency listener. Both are from the same lines of Story C, *The Indian Man*.

Excerpt W: SC in NNS speech to listeners (ES)

NNS8 (WLT): speech to ES listener (Story C)

> **S:** /*while* he :: looked around for someone to :: help him . a little boy :: walked
>
> up to him and :: started talking to him/ (4/1) .. /the old man- the man- the
>
> Indian man :: talked to the boy/ (1/1) /and *while* they :: spoke another man
>
> with dark glasses :: came up from maybe behind him .. :: picked up the basket
>
> and :: ran/ (4/1)... /the boy :: ran with the man *who* :: stole the basket/(2/1) /.
>
> and then the Indian man :: ran to the police officer/(1/1) .. /the policeman
>
> :: blew his whistle but by the time . they :: tried to :: stop them the man and the
>
> boy *who* :: stole this basket .. :: were in a car and they :: ran away/ (6/1)...
>
> /they :: drove really fast/(1/1)
>
> > [lines 5-16]

Excerpt X: SC in NNS speech to listeners (low)

NNS8 (WLT): speech to low-proficiency listener (Story C)

> **S:** /... a little boy . :: came to the Indian man .. and :: started to :: talk to
>
> him/(3/1) ... /now . as he :: spoke to him . there :: was another man *who*
>
> :: came . with dark glasses and :: stole the square basket /(4/1) .../ the man
>
> :: ran away and so :: did the boy/(2/1) ... /*while* they :: ran . the Indian man .
>
> :: ran to a police officer with a whistle/(2/1) .../ the police officer :: blew his
>
> whistle. . but . the man and the boy . :: got away ... /(2/1) /they :: were in car/
>
> (1/1) /and they drove- :: drove away very fast/ (1/1)
>
> > [lines 6-15]

Although both excerpts were taken from the same NNS, her speech was
adjusted for the different listeners. While the excerpt to the ES included a total of 19
S-nodes and seven clausal utterances (SC ratio: 2.70), the excerpt to the low-

proficiency listener included only 15 S-nodes and seven clausal utterances (SC ratio:

2.14). Therefore, while the NNS's speech to the ES used more syntactically complex

utterances, her speech to the low-proficiency listener mostly incorporated shorter,

simpler clauses. While there were a few complex utterances in excerpt X (e.g., /now

as he :: spoke to him . there :: was another man *who* :: came . with dark glasses and

:: stole the square basket/) and a few simple utterances in excerpt W (e.g., /the old

man- the man- the Indian man :: talked to the boy/), the overall difference between the

two excerpts suggests that the speaker purposefully modified her input to use less

complex syntax with the non-native listener.

For instance, in one segment, where the *policeman was trying to catch the*

thieves, the speaker used two distinct utterances in her speech to ES and a Low

proficiency NNS. In her speech to the low-proficiency listener, she simply said, "/ the

police officer :: blew his whistle. . but . the man and the boy . :: got away ... /(1/2);"

however, in her speech to the native listener, she said, "/the policeman :: blew his

whistle but by the time . they :: tried to :: stop them the man and the boy *who* :: stole

this basket .. :: were in a car and they :: ran away/ (6/1)." The former utterance

included few S-nodes per clausal utterance and was very straightforward, while the

latter included more S-nodes per clausal utterance and mentioned more details.

Interestingly, in an earlier example (excerpt T), another NNS elaborated on

his utterance to a non-native listener (e.g., referencing who the thieves were). This

reflects variance within speaker type due to individual differences – every speaker has

a different style of input. However, examining the example from excerpt T for

syntactic complexity reveals that the utterance "/ they :: rode in a car /(1/1).. /the man

with the sunglasses and the boy . they rode- they :: got into the car . and they

dro- :: drove /(3/1)" has four S-nodes and two casual utterances – less complex than

the example from excerpt X. This means that although both speakers clarified *thieves*, one used complex syntax (excerpt X, for her speech to a native listener) and the other used simple syntax (excerpt T, in his speech to a non-native listener). These examples illustrate the extent to which one NNS altered her speech through use of simpler, shorter utterances with a low-proficiency listener and more complex, longer utterances with a native listener. These findings for NNSs in terms of syntactic complexity were similar to earlier results found in NSs.

5.4.4 Content dilution

Finally, the last hypothesis under RQ4 (H4d) stated, *NNSs will omit more content details in their speech with low-proficiency listeners. This difference will be indicated by lower counts of IBs mentioned for low-proficiency listeners.* The NNS-only model found no significant difference in NNSs' speech between ES and Low or between ES and Hi. The overall results, however, with both speaker conditions in one analysis, showed a significant difference between ES in comparison to Hi and Low, indicating that a possible lack of power in the separate analyses may have prevented a statistically significant difference from being detected. Nevertheless, descriptively, the results suggest that advanced NNSs were inclined to omit more IBs when speaking to low-proficiency listeners than when speaking to the native controls. That is, low-proficiency listeners received slightly fewer content details (*M: 18.36 - SD: 3.91*) in comparison to native controls (*M: 19.13 - SD: 3.24*); for high-proficiency listeners, as it had been observed, the count of mentioned IBs was in between the other two listener types (*M: 18.43 - SD: 3.24*) and was not statistically significant.

Although the non-significant results appeared in the analysis of NNSs only, the overall results (both main model and descriptively) suggest that NNSs tended to mention fewer propositions when speaking to low-proficiency listeners, leading to content dilution. Therefore, further illustrations using examples from NNSs'

transcripts, along with additional interpretations, were provided. Excluded IBs were similar to those found in NSs' transcripts, but they varied from speaker to speaker. A NS and a NNS might both omit mentioning that "the Indian man arrived at the airport" and might include that "the basket was heavy," each in his or her own style. Each was compared within his or her own speech to native and non-native listeners (as discussed in using the type of random intercept in LME). These examples and others were compared within NNSs and across listeners. To demonstrate the type of differences found in IBs between ES and Low, the following two excerpts are reproduced from the same NNS's speech to an ES and to a low-proficiency listener. Both are from the same lines of Story A, *The Elephant Weight*.

Excerpt Y: IB in NNS speech to listeners (ES) (I)

NNS4 (WLT): speech to ES listener (Story A)

> **S:** So one day (1) :: at a port .. (2):: some officials .. men who were (3):: trying to weigh . (4)::an elephant .. and (5):: the only scale they had was (6)::.. not enough to fit the elephant . on ... so what they did .. a little boy was around . (7):: he suggested (8):: they . put the elephant on a barge .. in the water and ... then they mark- (9):: once the elephant is on they (10):: mark the barge .. (11):: where the level of water and then they take- they would (12):: take the elephant out ... and then (13):: fill . the barge with rocks .. until (14):: it sinks to that level of the elephant . with the elephant weights- because there- .. (15):: to make equivalent . and then they would remove the rocks and (16):: place it on the scale in (17):: a bucket ... and (18):: through that they (19) :: will be able to . weigh the elephant ..
>
> [lines 1-16]

Excerpt Z: IB in NS speech to listeners (low) (I)

NNS4 (WLT): speech to low-proficiency (Story A)

> **S:** Okay there were some . people (1):: at the port . (2):: trying to weigh (3)::an elephant ... the scale they had was (4):: too small . for the elephant ... so (5):: a little boy . suggested . that (6):: they . place the elephant. on a . barge .. and the barge- ... and then they (7):: mark . the barge . or (8):: the water line . where the barge sinks in ... then (9):: they remove . the elephant and (10):: fill the barge with rocks .. and- . (11):: until the barge sinks in . the water . to the line the- . the weight of the elephant .. and then they remove (12):: the rocks . into buckets and (13):: they weigh the rocks . (14):: to figure out . the weight of the elephant ..

> [lines 1-12]

Both excerpts were segmented to mark the propositions (IBs) counted, and were numbered for the purpose of describing the missing information. While the excerpt to ES included 19 information bits (story total: 19), the excerpt to the low-proficiency listener included only 14 (Story total: 14). In narrating the story to the low-proficiency listener, the NNS modified her input by omitting propositions such as *official men, the only scale available*, and *the elephant on board* (propositions 2, 5, and 9 in Excerpt Y). These IBs were mentioned and elaborated on in her speech to the native listener, indicating content dilution caused by the level of input modification she used with the low-proficiency listener. Two more excerpts from another NNS are given below to show other ways in which content was diluted.

Excerpt AA: IB in NNS speech to listeners (ES) (II)

NNS9 (NLT): speech to ES listener (Story B)

S: And then all of the sudden . (1)::a woman . came out of the car . and (2)::

slammed the door ... Okay .. after that .. (3)::the young kid came and placed

the coins (4)::into the can . that was in the hand of the blind man ...okay .. and

then all of the sudden ... the- the- the blind man was ex- expressing gratitude

to the young woman . but not the kid ... yes . so he had- he had like- he had a

hat and so ... as a courtesy he- (5)::he took his ha- hat and like wave it (6):: to

the young lady . (7)::as a way of thanking her ..

L: Why did he thank her not the kid?

S: Because he slammed the door- she slammed the door . so (8)::he was

thinking . (9)::that the lady .. was the one who gave them- gave him the money

      [lines 19-39]

Excerpt BB: IB in NS speech to listeners (low) (II)

NNS9 (NLT): speech to low-proficiency (Story B)

S: there was a- (1)::a woman . that stopped the car . and got out of the car okay

L: And what?

S: Got out of the car .. so she- she stopped- she- she stopped and then she got

out- outside the car ... and she when she got outside the car . (2)::she slammed

the door .. she closed the door very hardly Okay? . making a big noise of .

when you close the door you know ...

L: yes

S: this- . so .. (3)::the same time- so when- when- when the wom- when the

woman closed that door . very hardly .. the- the- (4)::the young boy came and

put the money . (5)::into the cup . of the old man ... okay. the old man . could

not see who gave him the money . because he was blind ... and so . instead of

. thanking the boy ... okay? . instead of . saying thank you to the boy ..

(6)::the- the- the . old man said thank you (7):: to the lady ...

[lines 59-79]

While the excerpt to the ES included nine IBs (story total: 24), the excerpt to the low-proficiency listener included only seven (story total: 19). In narrating the story to the low-proficiency listener, the speaker modified his input by omitting propositions such as *the blind man took his hat off, the blind man thought or assumed*, and *the blind man thought the lady was the one who gave him the money* (propositions 5, 8, and 9 in Excerpt AA). Interestingly, the NNS mentioned some of these details to the native listener based on her feedback. The native listener asked, "Why did he [the blind man] thank her not the kid?" Her question prompted the speaker to include important details he had not mentioned previously. This example showcases the types of interaction that can occur when the listener understands the content of the story sufficiently to clarify any information gaps in response to the right question. It is possible that this only happens when the speaker and the listener have similar levels of proficiency.

On the other hand, the low-proficiency listener requested a non-specific clarification ("and what?"), leading the speaker to repeat what he had said rather than add new content information. It is understandable, as the low-proficiency listener's main concern was to comprehend lexically what the speaker was saying rather than the sequence of the story. However, this leaves the listener trying to fill multiple comprehension gaps with his own interpretations; this shortcoming was usually revealed in the listener content comprehension measure. In other words, even with input modification, listeners with low proficiency often become disoriented when new concepts are introduced, especially when the details are incomplete.

The general results indicated that NNSs followed a similar pattern of content dilution to that found for NSs. That is, native listeners received significantly more IBs from NNSs than did low-proficiency listeners. These results demonstrated not only the occurrence of input modification, but also suggested that NNSs diluted content more when speaking to low-proficiency listeners than to NS listeners. However, as discussed in section 5.1, NNSs were inclined to retain more IBs in their speech to non-native listeners than were NSs, who mentioned significantly fewer IBs. This raises an interesting question for future research as to the reasons the NNSs retained more content information than NSs in their speech to non-native listeners.

5.4.5 Summary

The results in this section attest that NNSs do modify their input when speaking to low-proficiency non-native listeners in terms of lexical diversity, lexical sophistication, and syntactic complexity (Table 32). The difference in NNSs' speech between ES and Hi listeners was statistically significant only in the syntactic complexity measure. The content measure indicated non-significant differences in content dilution across the listener groups.

Linguistic and content measures in NNSs' speech				
Listener Type	CTTR	CVS1	SC	IB
ES	4.03	0.27	2.30	19.13
HI	3.91	0.22	2.04*	18.43
LOW	3.84*	0.19*	1.85*	18.36

Asterisks represent statistically significant differences in comparison to ES

Table 32: Linguistic and content measures in NNSs' speech

5.5 Listeners' content comprehension

All of the above results were discussed in terms of the speakers' input, using two process variables: linguistic and content. The findings described the input modification exhibited by speakers, both NSs and NNSs, in the three listener conditions. This section addresses the final research question (RQ5): the extent to

which input modification (especially content dilution) influenced listeners'

performance on the comprehension assessment (LC).

5.5.1 LC scores between high- and low-proficiency listeners
Since the non-native listeners' content comprehension assessments were in

Arabic, these two levels were compared separately first. The first hypothesis under

RQ5 (H5a) stated, *high-proficiency listeners will exhibit greater comprehension of the*

stories than low-proficiency listeners. This difference will be indicated by higher

content comprehension assessment scores. The model found a significant difference

in LC scores between high- and low-proficiency listeners. The low-proficiency

listeners scored statistically significantly worse (*M: 6.31 - SD: 4.48.38*) than the high-

proficiency listeners (*M: 10.91 - SD: 4.20*). The results showed that proficiency level

played a key role in determining non-native listeners' content comprehension.

5.5.2 LC scores between native and low-proficiency listeners
Another comparison was conducted to examine whether native listeners had

higher comprehension scores than non-native listeners. The second hypothesis under

RQ5 (H5b) stated, *ESs listeners will exhibit greater comprehension of the content*

than both low- and high-proficiency listeners. This difference will be indicated by

higher content comprehension assessment scores. The model found significant

differences in LC scores between ES and high-proficiency listeners and between ES

and low-proficiency listeners. Because they learned the content of the stories through

their native language, ES listeners comprehended more of the content (*M: 14.48 - SD:*

4.17) than both high- and low-proficiency listeners (whose means were stated above).

This comparison suggests, unsurprisingly, that using the native language in narrating

the content yields a better understanding of content details. It is also likely that with

increased proficiency level, comprehension of content also increases.

Furthermore, there was an interaction between Speaker Type and Listener Type. Interestingly, the high- and low-proficiency listeners in the NNSs' condition performed better in LC than the high- and low-proficiency groups in the NSs' condition. It is possible that NNSs mentioned the IBs in ways that were more salient to non-native listeners NSs. Further research is needed. Notably, as discussed in section 5.1.4, NNSs included more IBs with low-proficiency listeners than did NSs. These results suggest a possible link between the number of IBs mentioned by speakers and listeners' comprehension scores. The fact that NNSs speakers mentioned more IBs to low-proficiency listeners may have led to higher scores on the comprehension assessment for those listeners than for their counterparts paired with NSs. On the other hand, native listeners paired with NSs scored better on LC than native listeners paired with NNSs, probably because the NSs mentioned more details to the native listeners than did the NNSs.

Finally, story difficulty influenced both the number of mentioned IBs and the scores on LC – i.e., Story A was perceived as difficult, causing fewer IBs and poorer LC scores – suggesting that content difficulty may affect the number of mentioned IBs. Likewise, each Speaker Type performed differently in narrating each story, indicating that the speakers' perceptions of content difficulty can also affect overall results.

5.5.3 Summary

The results in this section show that non-native listeners overall scored worse than native listeners on the LC assessment. In addition, high-proficiency listeners significantly outperformed low-proficiency listeners. This suggests that language proficiency played a key role in listeners' understanding of the stories' content. Using a non-native language may adversely affect content learning by intermediate or low-proficiency learners. Content difficulty may also affect listeners' performance, as well

as the degree of input modification employed by speakers. Interestingly, an interaction between Speaker Type and Listener Type was found. Namely, the high- and low-proficiency listeners paired with NNSs performed better in LC than the high- and low-proficiency listeners paired with NSs.

Chapter 6: Conclusions and Implications for EMI

This study investigated input modifications in using English to narrate stories to learners of varied English-proficiency levels and the extent to which such modifications resulted in content dilution. It provided evidence of input modifications in a laboratory experiment that was meant to simulate some aspects of content-and-language-integrated settings, such as EMI, rather than purely language-learning contexts. Input modification was, and still is, viewed as positive in interaction between learners and instructors, as it allows students to receive input that is adjusted appropriately for their level, i.e., comprehensible input (Krashen, 1981, 1982), which enables them to build proficiency. However, in the context of content-and-language-integrated learning and EMI, this perspective on input modification may be inaccurate.

Much earlier research explored input modification by NSs, and only limited research has shown that NNSs modify their speech when addressing non-native learners. The results obtained in this study corroborated earlier findings regarding NSs' lexical and syntactical input modification when geared towards low-proficiency, non-native speakers. The findings here also confirmed that NNSs exhibit similar patterns in simplifying their speech to non-native speakers, although their input modifications differ from NSs' in terms of syntactic complexity and lexical sophistication. While NSs' and NNSs' speech only differ in lexical sophistication and syntactic complexity, providing the speakers with the three stories scripts, in their preparation for the study session, may have reduced a larger difference between NSs and NNSs, particularly in content dilution. However, the significant effect of speaker type on lexical sophistication and syntactic complexity did not have a substantial effect on the listeners' content learning. That is, listeners' content comprehension

scores were similar regardless of whether they were paired with NSs or NNSs; language native-ness did not significantly affect the way the stories were narrated.

In addition to linguistic modification, NSs omitted significantly more information bits when speaking to non-native listeners – evidence of content dilution as a result of input modification. NNSs exhibited no significant difference in their speech to native and non-native listeners, although the difference approached significance and descriptively showed a similar pattern to NSs. NSs, however, included *fewer* IBs in their speech to low-proficiency listeners than did NNSs.

The results indicated that language teaching (LT) experience did not have a significant effect on the way input and content were delivered. While there was a significant effect of LT experience on lexical diversity, it did not have a major impact on the content learning; the LC scores for listeners paired with speakers WLT were similar to the LC scores for listeners paired with speakers NLT. As reported above, speakers WLT responded to listeners by giving more feedback and attending to their questions, but it did not result in higher LC scores for those listeners. Although this style of teaching seems effective, it may not improve the learning outcome in content classrooms where students have a very poor command of English. In fact, the most significant predictor of LC score was listener proficiency. This suggests that, in order for students to comprehend the subject matter, they should have sufficient proficiency in the language. Instructors might modify their input through various means in order to improve students' comprehension, but doing so may not necessarily improve subject matter learning. Students must have reached a certain level in English in order to benefit fully from their instructors' input and feedback.

The majority of instructors in EMI contexts are intermediate or advanced NNSs of English; therefore, the input modification seen in the NNSs' speech in this

laboratory study may be similar to the types of input modification used in EMI classrooms. Although not directly generalizable, considering the possibility that these types of modifications may change over time in a genuine classroom setting, these findings may give researchers a starting point for understanding the *quality* of input provided by NNS instructors in content-and-language-integrated settings. These results demonstrated quantitatively and qualitatively that NNSs' input to low-proficiency learners tended to be downsized in several ways. Of course, further studies are required to explore NNSs' input modification in genuine EMI contexts.

One of this study's main findings was the key role played by listener proficiency in listener comprehension. Listeners with low proficiency tended to receive the least rich input in terms of lexical diversity, lexical sophistication, syntactic complexity, and informational content. These two mutually reinforcing findings may have implications for EMI contexts. EMI's central purpose is to deliver content through English; however, as discussed in the literature review, most students in EMI settings have a poor command of English. This leads instructors to develop methods to "cope" with the students' low-proficiency (Al-Kahtany et al., 2016, Shamim et al., 2016; Suliman & Tadros, 2011). The results of this study may offer new insights into the findings of previous studies of EMI contexts. Al-Kahtany et al. (2016), for instance, reported that undergraduate students faced difficulties in understanding content taught in English, as reflected by their overall academic achievement (60% of the students' GPAs were 1-1.99 out of 5). In addition to the students' poor command of English, findings reported in this study suggest that their low achievement may have been aggravated by input modifications. Because of their students' low English proficiency, the instructors may have simplified their input, thereby diluting the content and omitting key information. Input modified in this way

may help learner understanding of *language*, but reduced language, and as in this study, it may hinder their comprehension of *content*. In content classrooms, specific vocabulary and complex utterances are often required to reflect the nuances of the subject matter. Certainly, students with poor command of English would struggle to learn content through a language barrier; however, this study further implies that the teachers themselves may react to their students' low proficiency by modifying their speech and diluting the subject matter. It is important to note that this study has not shown which features of input, isolated, lexical, syntactic or content, affect students' learning; rather input modifications in general influenced students' learning of the content.

The goal of this research was to examine the types of input modifications employed by both NSs and NNSs when speaking to listeners of varying proficiency levels. Although the one-on-one experimental design used in this study cannot be generalized to an EMI classroom, where a single instructor has to deal with a whole group of students, longitudinally, the results suggest that both speaker types do modify their input according to listener proficiency. These types of input modifications may also be reflected in teachers' speech, in the classroom, when delivering content instruction through English. Based on the findings discussed above, this study suggests three new directions for EMI that might contribute to more effective content learning: (1) use the L1 – in this case, Arabic – as the medium of instruction for teaching subject matter (Arabization); (2) increase students' English proficiency; or (3) train content instructors in how to modify their speech effectively, including how to provide elaborated rather than simplified input.

6.1 Arabization

The most obvious implication from this study is to use students' L1 to teach content, particularly in specialized fields. The native English listeners in this study

statistically significantly outperformed both high- and low-proficiency non-native listeners on the content comprehension assessment. Even though, the high- and low-proficiency listeners paired with NNSs performed better in LC than the high- and low-proficiency listeners paired with NSs, their scores were still statistically significantly lower than the native controls, ES. Thus, using the students' L1 will considerably improve their content comprehension. Using a non-native language as a medium of instruction, as illustrated in this study, may not have a content-learning advantage. On the contrary, EMI and language-and-content-integrated learning may pose major obstacles to students' learning and overall achievement (Al-Kahtany et al., 2016). Using the students' native language would resolve this issue.

Beyond the potential shortcomings suggested here, learning content through another language can also result in discrepancies between theoretical knowledge and practical application. For instance, medical students in the Arab world often learn in English and then practice in Arabic (Al-Kahtany et al., 2016; Sabbour, Dewedar, & Kandil, 2010). Sabbour et al. (2010) found that this practice led to inadequate Arabic medical communication skills: medical students struggled with doctor-patient communication (e.g., taking medical history, explaining conditions and/or tests required, and answering patients' questions). Harmsen, Bernsen, Bruijnzeels, & Meeuwesen (2008) suggested, moreover, that medical students need specific training in the various regional dialects of patients throughout the country in an attempt to bridge potential cultural and language barriers. For doctors, these communication skills are essential and should be taught in the language of use.

In light of such findings, a number of studies conducted in the Arab world have argued for using Arabic to teach scientific content: a trend called "Arabization" (Al-Shbiel, 2017; Sabbour et al., 2010; Suliman, 2004). Arabization does not mean

forsaking English; "rather the aim is for the students to learn better in their native language while maintaining a good knowledge of English as the leading language of medical research today" (Sabbour et al., 2010, p. 1264). While the implications seem sensible and straightforward, they are challenging to implement. The limited number of academic resources in Arabic constrains its use as the medium of instruction in scientific fields. An interested university could, however, resolve this by investing in its professors and researchers to establish and publish their own scientific Arabic-language textbooks in their own fields. This important initiative could also be financially supported by the country's government, which would elevate the status of Arabic worldwide and strengthen its use in academia.

One rationale for using EMI in scientific fields is that much of the relevant academic knowledge worldwide is disseminated in English; therefore, students need to be competent in English, so they can access and use those resources (Galloway et al., 2017). While this rationale is understandable, the way EMI is currently implemented may not contribute to students' language or content learning. Learning a second language requires pedagogical methods distinct from those for content learning; students may not become proficient in English simply because they are in CLIL or EMI classrooms. Instead, a university could offer specific English courses for scientific and medical students, such as English for Specific Purposes (ESP), which should be taught using modern approaches, like Task-Based Language Teaching (TBLT). In this way, students could boost their English proficiency in order to read about global advances in their fields of study, as well as effectively learn and apply the content through their native language.

Worth noting, studies on bilingual programs, established in the US, raised parallel conclusions where immigrant students in primary and secondary schools

generally prefer the use of their native language—e.g., Spanish – for content courses. Studies in these contexts favored the use of a bilingual approach (classroom instruction with the L1 used before or simultaneously with the L2) which showed an advantage over a structured English immersion (SEI) approach (classroom instruction in English only). Overall, research showed that bilingual education supports the students' L1 as they develop their L2, promotes students' academic achievement, and creates positive learning environment (See, Greene, 1998; MacSwan, Thompson, Rolstad, McAlister, & Lobo, 2017; McField & McField, 2014; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005). Thus, the approach of teaching content through the students' L1, while simultaneously learning L2 in an ESP setting, provides the best possible learning outcomes, for both content and language.

6.2 Increase students' English proficiency

If using the students' native language is not possible, the second option is to improve students' English proficiency. It is common sense that students with higher English proficiency will, in general, be more successful in learning content delivered in English; certainly, this was supported in this study, which found that high-proficiency listeners outperformed low-proficiency listeners on the content comprehension assessment. Improving students' English proficiency, however, requires providing students with adequate instruction and practice, most likely in a foreign language context. In Saudi Arabia, current public EFL education would have to be reviewed thoroughly, closely examining the reasons it fails to equip students with sufficient language proficiency for English-medium higher education.

For students planning to enter higher education, a series of ESP courses should be developed. These could provide students with the necessary English terms and practice relevant to their own fields, such as instruction in reading and writing scientific journal articles. Students should be required to take these types of courses

either before or simultaneously with content courses. Of course, implementing this series of courses would involve several major steps: (1) conducting a needs analysis of the academic field and its essential knowledge (as in TBLT; Long, 2015), (2) course development by a team of language and content experts, (3) implementing pilot versions of these courses, (4) evaluating the courses' success, based on assessments of students' proficiency, and (5) further modification of the courses as needed.

6.3 Elaborated input

The third and final option to support effective content-learning through EMI is to use elaborated, rather than simplified, input. Input elaboration is "intended to help learners gain access to the meaning of [a] text by adding grammatical phrases and clauses such as defining appositives, relative clauses, and restatements" (Chapelle, 2003, p. 51). In an EMI classroom with learners of lower proficiency, elaborated input can preserve the lexis and subject matter information necessary for content learning, while clarifying its meaning and thereby supporting language acquisition. Instructors can use elaborative phrases, repetition and many other methods, or emphasize key information by stressing a word or even adding a one-beat pause before and/or after introducing key lexical items (Long, 2015).

Several SLA studies have found elaborated input to be more effective than simplified input. For instance, Oh (2001) showed that while simplified input only worked well for advanced L2 learners, elaborated input had a more consistent effect on comprehension across proficiency levels. Parker and Chaudron (1987) reviewed a number of studies comparing the effects of input modification (simplified and elaborated) and found that elaborative modifications had a consistently superior effect. Earlier research also suggested that elaborated input improved vocabulary learning in modified reading materials (Hulstijn, 1992; Kim, 2003, 2006). Elaboration

essentially provides L2 learners with semantic details, helping them draw accurate inferences from texts (Yano, Long, & Ross, 1993) or spoken discourse, thereby, improving their comprehension. While there is limited research on input elaboration in speech and its influence on comprehension, the above-mentioned findings offer compelling evidence for EMI classrooms to incorporate elaborated input. Elaborated input seems particularly well-suited to EMI, because it can provide learners with input enriched with unfamiliar words and their meanings, rather than omitting them altogether. However, it is not clear how much low-proficiency learners, such as those in this study, would benefit from elaborated input in learning the content.

Furthermore, EMI content instructors are often not trained language teachers. Any input modification they employ is simply a natural response to their students' perceived proficiency level; in fact, in this study, NSs and NNSs with and without language teaching experience employed, to varying degrees, similar modifications as shown by all measures, with the exception of CTTR. Whereas further investigation is required to warrant implementing elaborated input in EMI contexts (such as its potential benefits for low-proficiency learners), a plausible implication is to train content instructors in input modification and emphasize elaborative techniques. These training programs would have three aims: (1) raise awareness of the input modifications instructors naturally make, (2) describe the difference between simplified and elaborated input, and (3) stress the importance of retaining as much of the content as possible.

6.4 Limitations and directions for further research

This study contributes a different perspective on input modification for the SLA literature. Generally, input modification has been recognized as being employed by NSs speaking to non-native listeners. In addition to corroborating this type of modification, the current study found that NNSs employed similar input modification

in their speech towards non-native listeners, as operationalized through measures of lexical diversity, lexical sophistication, and syntactic complexity. Furthermore, analysis of propositions mentioned in both NSs' and NNSs' speech to non-native listeners revealed significant content dilution in contrast to their speech to native listeners. While these results are not generalizable to an EMI context because the study was conducted in a laboratory setting, they do reveal, in a rigorous and controlled situation, the type of input modification that other NNSs – such as instructors in EMI classrooms – may exhibit. In order to generalize these findings, the operationalized concept of content dilution, as well as measures of lexical and syntactic complexity, should be explored further in different settings, such as genuine EMI contexts, and through different experimental designs. Further research should also incorporate a longitudinal design, along with multiple, longer content lessons (e.g., science) taught to several low-proficiency students at a time (not one-on-one). In addition to examining the teacher's speech, future studies should include both vocabulary and subject matter tests to measure the learning that takes place.

Notably, the results indicate that both content (level of difficulty) and speaker (background, style, and individual differences) are moderating variables that affected the outcome of the study. This was the rationale behind using Story-Speaker as a random intercept in LME models, which in turn reflected the distinctive dependency structure established in the experimental design. Although the use of multiple stories in this study was mainly to preclude the confounding influence that only one might have, the differences between stories affected, to some extent, the input modification employed by speakers. Depending on story difficulty, the results from analyses of the linguistic and content measures varied in significance. In general, it appeared that when the content was perceived as difficult or contained more complex technical

details or abstract notions, the more input was modified and the fewer content details were mentioned; however, because content difficulty was not controlled for in this study and participants were not asked about their perceptions of each story's difficulty, this interpretation is only conjecture. It was beyond the scope of this study to explore, using empirical methods, differences in quality and complexity between the stories, let alone establish a measure to assess the difficulty of various subject matters. Future research should explore the influence of content lesson difficulty on both the number of content details mentioned in instructors' speech and the students' overall content learning. In order to control for content difficulty, researchers should develop a systematic and valid measure to define what constitutes difficult or complex content lessons in comparison to easy or simple content lessons. Systematically defining content difficulty will, in turn, provide a better understanding of how difficulty determines the number of information bits mentioned in content classrooms.

This study's findings suggest that listeners' perceived proficiency directly affected the extent of the observed input modification. Across all measures used in this study, the level of linguistic and content complexity increased with listeners' proficiency levels. Therefore, future research should investigate to what extent proficiency level triggers input modification and content dilution by using participants of varying proficiency. This type of research will assist in determining the appropriate level required for students to effectively learn content through EMI.

As mentioned in the discussion of implications for EMI, future research should also explore elaborated input and its ability to deliver and preserve key content. Whereas much research has explored the effectiveness of elaborated input in written discourse, further research is needed on its use in spoken discourse and its potential influence on content and vocabulary learning. Although EMI's goal is not to

teach language, elaborated input could assist students in learning key vocabulary items in content courses.

Finally, this study found that although language teaching experience qualitatively influenced the types of interactions between speaker and listener (e.g., responding to a listener's cues and providing necessary feedback), it only found significant differences between the two groups in lexical diversity. This difference, however, did not influence listeners' content learning, possibly because their low English proficiency prevented them from benefiting from this type of interaction. Future research should explore the extent to which language teaching experience influences language and content learning in L2 learners with higher proficiency levels. High-proficiency students may benefit from the kind of input delivered by instructors with LT experience compared with instructors without LT experience.

Appendix A: Dictation test (English listening proficiency measure)

Name: _____   **Date: _____**

Dictation test

Directions: You will hear a short story twice. The first time, just listen to the whole story (about a minute). The second time, each sentence will be read again, separately, followed by a pause for you to write down what you hear. Write as much of what you hear as possible. Write carefully and legibly.

التوجيهات: سوف تستمع لقصة قصيرة مرتين. في المرة الاولى فقط استمع للقصة بأكملها (لمدة دقيقة تقريبا). وفي المرة الثانية سوف تستمع لكل جملة على حدة, مع وقفة بين كل جملة لكي تكتب ما تسمع. يرجى كتابة أكبر قدر ممكن من الكلمات والجمل المسموعة على هذه الورقة باللغة الإنجليزية. أكتب بحرص وبخط واضح.

Appendix B: Dictation test (Script)

Dictation script:

One day last week, Mark went grocery shopping. It took him a long time to do all his shopping, so he caught a bus home late in the evening. He had to walk home through a lonely forest and he felt a little frightened. Suddenly, he heard a noise. There was a strange man walking behind him. Mark started to walk much faster. The man behind him began to pick up his pace, too. The next moment the man caught up with him and grabbed his arm. The man said that he had followed Mark to give him a parcel.

With pauses:

One day last week, Mark went grocery shopping. PAUSE (8 sec) It took him a long time to do all his shopping, PAUSE (10 sec) so he caught a bus home late in the evening. PAUSE (9 sec) He had to walk home through a lonely forest PAUSE (8 sec) and he felt a little frightened. PAUSE (5 sec) Suddenly, he heard a noise. PAUSE (5 sec) There was a strange man walking behind him. PAUSE (8 sec) Mark started to walk much faster. PAUSE (5 sec) The man behind him began to pick up his pace, too. PAUSE (10 sec) The next moment the man caught up with him and grabbed his arm. PAUSE (12 sec) The man said that he had followed Mark to give him a parcel.

Appendix C: Picture-sequence sample (warm-up story)

The fisherman

Appendix D: Scripts for the main stories

* Adapted from Lynch (1987) and Heaton (1966)

Story A: The elephant weight

The story was set in a port where a group of officials were baffled by the problem of having to weigh an elephant that is too large for the only available scale. A small boy had an idea. He told them to put the elephant onto a barge. Then, the boy marked a line to show the water level with the elephant on board. Two men then took the elephant out and filled the barge with stones until it sunk back down to the line, when the weight of the elephant and the stones were the same. They were then able to use the scale to weigh the stones in the baskets. Based on the weight of the stones, the the boy calculated the weight of the elephant.

Story B: The blind man

The story started with a young boy looking into a toyshop window. He was deciding what to buy with his pocket money. On the other side of the street, he noticed a blind man who was a beggar. The boy suffered pangs of conscience and decided to donate his pocket money to the man, rather than spend it on himself. As the boy approached the blind man, a woman got out of her car and slammed the door. The man heard the noise just as the boy dropped his money into his collecting tin, so the man assumed that the car driver was the donor and took off his hat in a gesture of gratitude towards where the noise came from. The boy was disappointed that his good deed has gone unrecognized.

Story C: The Indian man

In the story, an Indian man with a basket arrived at an airport. The basket was too heavy to carry, so he put it down and began to look for a porter. Just then he noticed a small boy. The boy came up to him and began talking to him. At the same time, a man with dark glasses picked up his basket. Then, the man and the boy ran away. The Indian hurried to a policeman with a whistle. The policeman tried to stop the thieves, but it was useless. The boy and the man got into a car and drove off. The car passed a zoo and then went towards a wood. When it reached the wood, the car stopped. The man and the boy got out. They opened the basket and, to their great surprise they saw a snake.

Appendix E: Listener comprehension (LC) assessment

What did you understand from the story? What was it about?
Please restate and list as many details as possible from each story you heard.[37]

Story 1:[38]

1. -

2. -

3. -

4. -

5. -

6. -

7. -

8. -

9. -

10. -

11. -

12. -

13. -

14. -

15. -

16. -

---

[37] An Arabic version was given to the non-native speakers and they were instructed to answer in Arabic.

[38] The same format, using the same question, was followed for the second and third stories.

Appendix F: List of propositions/information bits

Each story has 26 propositions/information bits (78 in total). There are some propositions that have two or three sub-details marked by an asterisk and a double colon that separates them.

Story A: Elephant weight

1. The story was set in a port (dock).
2. There was an elephant
3. There was a group of officials (official/government men)
4. They were baffled (confused).
5. They had a problem (issue).
6. A problem of weighing the elephant (want to weigh they elephant).
7. The elephant is too large (big/doesn't fit).
8. :: There is one small scale :: available (only one they had).* **(2)**
9. A small boy had an idea (a suggestion).
10. The boy told officials (them/men) :: to put the elephant onto a barge (boat).* **(2)**
11. :: The boy :: marked a line. * **(2)**
12. :: The line is to show the water level (line) with :: the elephant on board.* **(2)**
13. The men then took the elephant out.
14. :: The men filled the barge with stones (rocks) until :: the barge sunk back down to the line (went down to the line).* **(2)**
15. The weight of the elephant and the stones were the same.
16. The men were then able to use the scale.
17. They weighed (measured) the stones :: in the baskets (buckets). **(2)**
18. :: Based on the weight of the stones, :: the boy :: calculated (figured out) the weight of the elephant.* **(3)**

Story B: The blind man

1. There is a young (little) boy.
2. :: He was looking into a toy shop (toy store) :: window.* **(2)**
3. He was deciding what to buy (what to get).
4. He has his pocket money (change/coins).
5. On the other side of (across) the street.
6. The boy noticed (saw) :: a blind man. * **(2)**
7. The blind man was a beggar.
8. The boy suffered pangs of conscience (felt guilty/bad).
9. The boy decided to donate (give) his pocket money to the man.
10. Instead of spending the money on himself (on toys).
11. :: As the boy approached (came/walked to) the blind man, :: a woman got out of her car.* **(2)**
12. The woman slammed the door (shut/closed the door hard/loud).
13. The man heard the door slam noise (sound).

14. :: At the same time (right as), the :: boy dropped (put) his money into :: man's tin (cup).* (3)
15. :: The blind man assumed (thought) that :: the car driver (woman) was the donor (who gave him the money).* (2)
16. :: The blind man took off (tips) his hat :: in a gesture of gratitude (say thank you).* (2)
17. The man gestured towards where the noise came from (towards the woman).
18. :: The boy was disappointed (sad) that :: his good deed has gone unrecognized (unnoticed/didn't know about).* (2)

Story C: The Indian man

1. In an airport.
2. An Indian man
3. He arrived (came in).
4. He has a basket (bag/luggage/suitcase).
5. :: The basket was too heavy to carry, so :: he put it down.* (2)
6. He began to look for a porter (someone to carry it).
7. He noticed (saw) a small boy.
8. :: The boy came up (walked up) to him and :: began talking to him.* (2)
9. There was also a man with dark glasses (sunglasses).
10. The man with dark glasses picked up (stole/took/picked up) the Indian man's basket.
11. :: Then the man and the boy :: ran away.* (2)
12. :: The Indian hurried (ran/told/yelled) to a policeman (police officer) :: with a whistle (blow his whistle/whistling).* (2)
13. :: The policeman could not stop (catch) :: the thieves.* (2)
14. The boy and the man got into a car and drove off.
15. The car passed a zoo (they drove past a zoo).
16. :: The car then stopped (arrived) :: at the woods (forest).* (2)
17. The man and the boy got out (jump out).
18. They opened the basket and, to their :: great surprise :: they saw a snake (there was a snake).* (3)

Appendix G: Instructions for speakers

During the session, you will sit face-to-face with one listener at a time. There are three main stages in the session:

1) Introduction and language level

- Get to know the listeners and their English language level by introducing yourself and asking them to introduce themselves (ask about their education, family, hobbies, etc.).
- Ask all of the following:
    - What are you going to do this weekend?
    - What do you do when you hang out with your friends?
    - What did you do on your last vacation?
    - What is your favorite food?
    - What is your favorite movie?
    - Which person from history would you most like to meet?

2) Warm-up stories

- Start narrating warm-up story one
- *Rules to follow:*
    - The story sheets will only be shown to you.
    - You will NOT be permitted to show the listener the pictures.
    - Talk about each frame in the story.
    - Nonverbal cues will NOT be permitted.
    - The listeners will depend on your narration of the story, so please make sure they understand it well.
    - The listeners are encouraged to ask questions and give you feedback about the story.
    - Pay attention to the listener feedback and respond accordingly
- After you finish telling the story, ask them to retell it to you.
    - If they understood it, move on to the second warm-up story.
    - If not, explain and correct the parts they got wrong.
        - You may repeat the story if necessary or upon request.
- Narrate the second warm-up story following the same rules and steps

3) Main stories

- Three main stories will be placed in front of you
- Narrate each story in the order it is given to you.
- *Follow the same rules as those in the warm-up stage with the exception of: :*
    - After you finish each story, the listeners will retell the story in their native language, writing on the paper provided.
    - Repetition of the stories will NOT be permitted
    - The listeners' understanding of each story will be tested in writing

Appendix H: Instructions for listeners

During the session, you will sit face-to-face with a speaker. There will be three main stages in this session:[39]

1) Introduction

- The instructor will get to know you by introducing himself or herself, and by you introducing yourself (talk about your education, family, hobbies, etc.).
- Answer his or her starter, icebreaker, questions

2) Warm-up stories

- The instructor will narrate the first warm-up story.
- *Rules to follow*:
    - You will not see the story content and depend on the speaker's narrations
    - You are encouraged to ask questions and request clarifications
- After the instructor finishes, restate the story.
- The instructor will then narrate the second warm-up story
    - Follow the same rules and steps

3) Main stories

- The instructor will be provided with three stories to narrate to you.
- *Follow the same rules as those in the warm-up stage with the exception of:*
    - After the instructor finishes each story, you will be asked to retell the story by writing it down on the paper provided. Be sure to include as many details as possible.

---

[39] An Arabic version was given to the non-native speakers.

Appendix I: Transcription guidelines

Format:
- Every word was transcribed as it was heard

- The speaker's speech is on the left side of the document and listener's speech is on the right side.

- (R): The researcher's speech in the listener's column.

- Each line is numbered

    - New stories start with new numbers

- Timestamp: The time each story ends in the audio file

Conventions:
- A period (.): one-beat pause

    - One-beat = one second pause

    - Two-beats (..), three-beats (…) and so on.

    - Fillers such as hmm, umm, uha were marked as pauses (…)

- A dash (-): A self-cutoff or interruption by another person

    - *He wants to bu- buy a toy*

    - *The woman slammed the doo- the car door*

- A question mark (?): A question

- Two question marks (??): Unclear/inaudible speech

- Two Xs (xx): Arabic speech with English translation in parentheses
- Italics: Emphasis

Appendix J: A sample transcript

NNS5 (NLT) ES11

	Speaker	Listener
1	**A little boy was** standing in front of a .. a toy store ..	
2	and he was looking at the window ... and as he was	
3	looking . he was thinking of . what he can get from his	
4	pocket change .. and then he saw a blind old man .. he is	
5	a beggar .. he was on the other side of the street ... so the	
6	boy felt bad so he decided that he will give the old man	
7	all of his money .. and as he was putting the change in	
8	the tin can that the old man was carrying ... a woman	
9	came out of the car and she slammed the door .. so the	
10	old man thought that the woman . actually . gave him the	
11	money and not the boy because he didn't hear where it	
12	was coming from .. and the boy felt bad because he didn't	
13	get the recognition .. of doing that something.	
14		
15	*Yes . That is it ...*	…
16	*[00:54]*	
	So the other story is at a seaport .. so seaport officials .	
1	they wanted to weigh an elephant .. but the scale they	
2	have does not fit him .. so . a smart boy came with an	
3	idea of using a barge .. and he suggested that they put	
4	the elephant on the barge .. and then once it sinks in the	
5	water . he marked where the water was .. then he took	
6	out the elephant ... and then decided that they can put	
7	stones and as much- they put stones and as much as they	
8	need to reach that mark he made .. then they took out the	
9	stones . weighed them . and he calculated the weight of	
10	the elephant.	
11		
12	Yes	Good . do I ask a question?
13		
14	It is like a boat .	… What is a barge?
15		
16		and what body of water,
17		like the ocean or like the
18	It said at a seaport .	pool
19		
20		Okay so . they are putting
21	On the barge- on the boat.	the elephant in the ocean
22		
23	Yes	Okay …
24	[02:00]	

1 2 3 4 5 6 7 8 9 10 11 12 13 14	**The third story is** at an airport .. so there was an Indian man .. he had- he was waiting at the airport . for a porter to come and help him with the basket he had it was very heavy and he could not carry it .. and as he was waiting . a young boy came and he was talking to him .. so he was distracted . talking to that boy .. another man who was wearing sunglasses came- and he was sneaking behind him .. he carried the basket .. and . that man and the boy ran- run off with the basket ... so the Indian man was screaming . and the police officer saw what is- what was happening . so he whistled for the thief to stop . but they didn't stop .. so they went on a car and drove off very fast .. they passed by a zoo .. and then decided that they would stop and check out what is inside the basket . and when they opened it . there was a snake . inside it.	
15 16 17 18 19 20 21 22 23 24 25 26 27 29 30 31 32	 No . So young person and the one … No… They are not wearing glasses .. just one man … [03:27]	Right . you say that . the other person was wearing glasses but you never mentioned that the first person is wearing glasses … Okay … So the young boy and the other person are wearing glasses … No . and why you said that they are also wearing glasses Okay … alright

Bibliography

Aburizaizah, S. (2015). Introducing content and language-integrated learning at a Saudi Arabian university. In R. Wilkinson, & M. L. Walsh (Eds.), *Integrating Content and Language in Higher Education: From Theory to Practice Selected papers from the 2013 ICLHE Conference* (pp. 101-116). Oxford: Peter Lang.

Admission Standards to Saudi Arabia Public Universities. (2016). In Riyadh Public Universities Admission Gate. Retrieved from https://www.rbu-admit.edu.sa/index.html#!/page/0011/1).

Aguilar, M., & Munoz, C. (2014). The effect of proficiency on CLIL benefits in Engineering students in Spain. *International Journal of Applied Linguistics*, *24*(1), 1-18.

Ai, H., & Lu, X. (2010, June). *A web-based system for automatic measurement of lexical complexity.* Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10). Amherst, MA.

Al-Kahtany, A., Faruk, S., & Al Zumor, A. (2016). English as the medium of instruction in Saudi higher education: Necessity or hegemony? *Journal of Language Teaching and Research, 7*(1), 49-58.

Al-Mohanna, A. (2010). English language teaching in Saudi Arabia context: How communicatively oriented is it? *Lang. & Transl., 22,* 69-88.

Al-Nofaie, H. (2010). The Attitudes of teachers and students towards using Arabic in EFL Classrooms in Saudi public schools: A Case Study. *Novitas-ROYAL (Research on Youth and Language), 4*(1), 64-95.

Al-Seghayer, K. (2005). Teaching English in the Kingdom of Saudi Arabia: Slowly but steadily changing. In G. Braine (Ed.). *Teaching English to the world: History, curriculum, and practice* (pp. 133-141). Mahwah, NJ: Lawrence Erlbaum.

Al-Shbiel, A. O. (2017). Arabization and its effect on the Arabic language. *Journal of Language Teaching and Research, 8*(3), 469–475.

Arnau, J. (2000). Catalan immersion teachers: Principles of language teaching. *International Journal of Bilingual Education and Bilingualism, 3*, 79–100.

Baayen, H., Davidson, D., & Bates, D. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(4).

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *Elementary School Journal, 83*, 177-181.

Belhiah, H., & Elhami, M. (2015). English as a medium of instruction in the Gulf: When students and teachers speak. *Language Policy, 14*, 3-23.

Brentjes, S. (2010). *Travellers from Europe in the Ottoman and Safavid empires, 16th-17th centuries: seeking, transforming, discarding knowledge*. Oxon, UK: Ashgate Publishing.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System, 39*(4), 523-32.

Bruton, A. (2013). CLIL: Some of the reasons why . . . . and why not. *System, 41*(3), 587-97.

Bruton, A. (2015). CLIL: Detail matters in the whole picture. More than a reply to J. Huttner and U. Smit (2014). *System, 53*, 119-28.

Bulté, B. & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds). *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam/Philadelphia: Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65.

Byun, K., Chu, H., Kim, M., Park, I., Kim, S., & Jung, J. (2011). English-medium teaching in Korean higher education: Policy debates and reality. *Higher Education, 62*, 431-449.

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice-Hall.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam/Philadelphia: Benjamins.

Cenoz, J. (Ed.). (2008). *Teaching through Basque: Achievements and challenges* (Vol. 21, No. 1). Bristol, England: Multilingual Matters.

Cenoz, J. (2014). Content-based instruction and content and language integrated learning: The same or different? *Language, Culture and Curriculum, 28*(1), 8-24.

Chang, Y. Y. (2010). English-medium instruction for subject courses in tertiary education: Reactions from Taiwanese undergraduate students. *Taiwan International ESP Journal, 2*(1), 55-84.

Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology* (Vol. 7). Amsterdam/Philadelphia: Benjamins.

Chapple, J. (2015). Teaching in English is not necessarily the teaching of English. *International Education Studies, 8*(3), 1–13.

Chaudron, C. (1982). Vocabulary elaboration in teachers' speech to L2 learners. *Studies in Second Language Acquisition, 4,* 170-180.

Chaudron, C. (1983a). Simplification of input: Topic reinstatements and their effects on L2 learners' recognition and recall. *TESOL Quarterly*, *17*(3), 437-458.

Chaudron, C. (1983b). Foreigner talk in the classroom – an aid to learning? In H. Seliger & M. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 127-143). Rowley, MA.: Newbury House.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* Cambridge, NY: Cambridge University Press.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564–570.

Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching, 39*(1), 1-14.

Corder, S. P. (1967). The significance of learners' errors. *IRAL, 5*, 161-9.

Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics, 11*(2), 183-199.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*(4), 66-79.

Cummins, J. (1998). Immersion education for the millennium: What have we learned from 30 years of research on second language immersion?. In M. R. Childs & R. M. Bostwick (Eds.) *Learning through two languages: Research and practice* (pp.34-47). Japan: Katoh Gakuen,

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369-382.

Dallinger, S., Jonkmann, K., Hollm, J., & Fiege, C. **(**2016**)**. The effect of content and

   language integrated learning on students' English and history competences – Killing

   two birds with one stone? *Learning and Instruction, 41*, 23-31.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL)*

   classrooms. Amsterdam/Philadelphia: Benjamins.

Dalton-Puffer, C. **(**2011**)**. Content-and-language integrated learning: From practice to

   principles? *Annual Review of Applied Linguistics, 31*, 182-204.

Dearden, J. (2014). *English as a medium of instruction- a growing global phenomenon*. The

   British Council ELTRA Reports (1–35). Retrieved from:

   http://www.britishcouncil.org/education/ihe/knowledge-centre/english-language-

   higher-education/report-english-medium-instruction

DeKeyser, R. M. (2000). The robustness of critical period effects in second language

   acquisition. *Studies in Second Language Acquisition, 22*(4), 499–533.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review

   of issues. *Language Learning, 55*, 1–25.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a stay

   abroad program. *Foreign Language Annals, 43*(1), 80–92.

Doiz, A., Lasagabaster, D., & Sierra, J. (Eds.) (2013). *English-medium instruction at*

   universities: Global challenges. Bristol, England: Multilingual Matters.

Drach, K. (1969). The language of the parent: A pilot study. In K. M. Drach, B.

   Kobashigawa, C. Pfuderer, & D. Slobin, *The structure of the linguistic input to*

   children (Working papers of the Language Behavior Research Laboratory, no. 14).

   Berkeley, CA: Language-Behavior Research Laboratory. Retrieved from:

   http://eric.ed.gov/?id=ED138109

Dudley–Evans, T., & Swales, J.M. (1980). Study modes in the Middle East. In the British

Council (Ed.), *Study modes and academic development of overseas students* (pp. 91-

101). London: British Council.

Early, M. (1985). *Input and interaction in content classrooms: Foreigner talk and teacher*

talk in classroom discourse (Unpublished doctoral dissertation). University of

California, Los Angeles.

Ellis, R. (2008). *The Study of Second Language Acquisition* (2nd ed.). Oxford: Oxford

University Press

Ferguson, C. (1971). Absence of copula and the notion of simplicity: A study of normal

speech, baby talk, foreigner talk, and pidgins. In D. Hymes (Ed.), *Pidgination and*

creolization of languages (pp. 141-150). New York: Cambridge University Press.

Fisher, M. C. (2011). *Dictation: What and how students learn from it* (Master's thesis).

School for international Training, Brattleboro, Vermont.

Floris, F. D. (2014). Learning Subject matter through English as the medium of instruction:

students' and teachers' perspectives. *Asian Englishes, 16*(1), 47-59.

Gaies, S. (1977).The nature of linguistic input in formal second language learning: Linguistic

and communicative strategies in ESL teachers' classroom language. In H.D. Brown,

C. Yorio, & R. Crymes (Eds.), *On TESOL'77*: *teaching and learning English as a*

second language: trends in research and practice (pp. 204-212). Washington,

DC.:TESOL

Gaies, S. (1980). T-unit analysis in second language research: Application, problems and

limitations. *TESOL quarterly, 14*(1), 53-60.

Galloway, N., Kriukow, J., & Numajiri, T. (2017). *Internationalisation, higher education and*

the growing demand for English: an investigation into the English medium of

instruction (EMI) movement in China and Japan. British Council ELT Research

Papers (Vol. 17). Retrieved from

https://www.teachingenglish.org.uk/sites/teacheng/files/H035 ELTRA

Internationalisation_HE_and the growing demand for English A4_FINAL_WEB.pdf

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Lawrence

Erlbaum.

Gass, S., & Selinker, L. (2001*). Second language acquisition: An introductory course* (2nd

ed.). Mahwah, NJ: Lawrence Erlbaum.

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual

education.* Cambridge, MA: Newbury House.

Goodman, B.A. (2014). Implementing English as a medium of instruction in a Ukrainian

university: Challenges, adjustments and opportunities. *International Journal of

Pedagogies and learning, 9*(2), 130-141.

Granowsky. S., & W. Krossner. (1970). Kindergarten teachers as models for children's

speech. *The Journal of Experimental Education, 38*, 23-8.

Greene, J. P. (1998). *A meta-analysis of the effectiveness of bilingual education.* Claremont,

CA: Thomas Rivera Policy Institute.

Hair, J. F., Black, W. C., Babin, W. C., & Anderson, R. E. (2010). Examining Your Data. In

Multivariate Data Analysis (pp. 31–88). Boston, MA.: Pearson.

Hakansson, G. (1986). Quantitative aspects of teacher talk. In G. Kasper (Ed.), *Learning,

teaching and communication in the foreign language classroom* (pp. 83-98). Aarhus,

Denmark: Aarhus University Press.

Hallett, R. W. (2000). *Simplified input: An investigation of foreigner talk/teacher talk on

comprehension and vocabulary acquisition (*Doctoral dissertation). Retrieved from

University Microfilms International (UMI). (No. 3006037).

Harley, B., & King, M. L. (1989). Verb lexis in the writ-ten compositions of young L2 learners. *Studies in Second Language Acquisition, 11*, 415–440.

Harmsen, J. A. M., Bernsen, R. M. D., Bruijnzeels, M. A., & Meeuwesen, L. (2008). Patients' evaluation of quality of care in general practice: What are the cultural and linguistic barriers? *Patient Education and Counseling, 72*(1), 155–162.

Heaton, J. B. (1966). *Beginning Composition Through Pictures.* United Kingdom: Longman Group.

Heaton, J. B. (1975). *Composition Through Pictures.* United Kingdom: Longman Group.

Henzl, V. (1973). Linguistic register of foreign language instruction. *Language Learning, 23*, 207-222.

Henzl, V. (1979). Foreigner talk in the classroom. *International Review of Applied Linguistics, 17*, 159-167.

Honeyfield, J. (1977). Simplification. *TESOL Quarterly, 11*(4), 431–40.

Hu, G., & Lei, J. (2014). English-medium instruction in Chinese higher education: A case study. *Higher Education, 67*, 551-567.

Hu, G., Li, L., & Lei, J. (2014). English-medium instruction at a Chinese university: Rhetoric and reality. *Language Policy, 13*, 21-40.

Hulstijn J.H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In Arnaud P.J.L. & Béjoint H. (Eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: Palgrave Macmillan.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs for the Society for Research in Child Development, 35*(1), 1-67.

Ishiguro, T. (1986). *Simplification and elaboration in foreign language teacher talk and its source* (Unpublished doctoral dissertation). Stanford University, Stanford.

Ivanova, J. (2011). *The effects of teacher talk on L2 learners' comprehension* (Unpublished master thesis). University of Utah, Salt Lake City.

Jackson, D. (2014). *The relative contribution of input modification, learner awareness, and individual differences to second language construction learning* (Doctoral dissertation). Retrieved from University Microfilms International (UMI). (No. 3582920).

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, *63*(1), 87-106.

Johnson, R. K., & Swain, M. (1997). *Immersion education: International perspectives.* Cambridge, UK: Cambridge University Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 20*, 60–99.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning, 42*, 217–248.

Kabouha, R., & Elyas, T. (2015). Aligning teaching and assessment to course objectives: The case of preparatory year English program at king Abdulaziz university. *International Journal Of Applied Linguistics And English Literature*, *4*(5), 72-81.

Kilickaya, F. (2008). Instructors' attitudes towards English-medium instruction in Turkey. *Humanising Language Teaching, 6*. Retrieved from: http://www.hltmag.co.uk/nov06/mart01.htm

Kleifgen, L. A. (1985). Skilled variation in a kindergarten teacher's use of foreigner talk. In S. M. Gass & C. G. Madden (Eds.)*, Input in second language acquisition* (pp. 59-68). Rowley, MA.: Newbury House.

Kim, Y. (2003). *Effect of input elaboration and enhancement on second language vocabulary acquisition through reading by Korean learners of English* (Unpublished doctoral dissertation). University of Hawai'I, Honolulu.

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly, 40*(2), 341–373.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.

Kuznetsova, A., P. B. Brockhoff, & B. Bojesen Christensen. (2015). lmerTest: Tests in Linear mixed effects models (Version 2.0-30), available at http://CRAN.R-project.org/package=lme4.

Kym, I., & Kym, M.H. (2014). Students' perceptions of EMI in Korea. *Journal of Asia TEFL, 11*(2), 35-51.

Landers, R.N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower*. DOI: 10.15200/winn.143518.81744

Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, *5*(1), 3-18.

Lasagabaster, D., & Sierra, J.M. (2009). Immersion and CLIL in English: More differences than similarities. *ELT Journal*, *64*, 367-75.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 307–322.

Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review*, *63*, 605-28.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Lei, J., & Hu, G. (2014). Is English-medium instruction effective in improving Chinese undergraduate students' English competence? *IRAL*, *52*(2), 99-126.

Li, N., Leung, D.Y.P., & Kember, D. (2001). Medium of instruction policy in Hong Kong universities: The mis-match between espoused theory and theory in use. *Higher Education Policy, 14*(4), 293-312.

Lo, Y. Y., & Lo, E. S. (2014). A meta-analysis of the effectiveness of English-medium education in Hong Kong. *Review of Educational Research*, *22*(10), 1-28.

Long, M. H. (1980). *Input, interaction, and second language acquisition* (Doctoral dissertation). Retrieved from University Microfilms International (UMI). (No. 8111249)

Long, M. H. (1983a). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition* 5, 2, 177-193.

Long, M. H. (1983b). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126–141.

Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass & C. G. Madden (Eds.). *Input in second language acquisition* (pp. 377–93). Rowley, MA: Newbury House.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Chichester: Wiley.

Long, M. H., & Ross, S. **(**1993**)**. Modifications that preserve language and content. In M. Tickoo (Ed.) *Simplification*: *Theory and application* (pp. 29-52). Singapore: SEAMEO Regional Language Centre.

Long, M. H., Al Thowaini, A., Al Thowaini, B., Lee, J., & Vafaee, P. (2018). A micro

    process-product study of a CLIL lesson: Linguistic modifications, content dilution,

    and vocabulary knowledge. *Instructed Second Language Acquisition*, *1*(2), 3-37.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral

    narratives. *The Modern Language Journal*, *96*(2), 190-208.

Lynch, T. (1987). Modifications to foreign listeners: the stories teachers tell. *Center for*

    Applied Linguistics. Retrieved from: http://eric.ed.gov/?id=ED274225

Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns

    across divergent contexts. *Language Teaching Research, 15*, 279–288.

Macaro. E. (2014). *EMI: Researching a global phenomenon*. Keynote speech presented at the

    British Council European Policy Dialogue, Segovia.

Macaro, E., Akincioglu, M., & Dearden, J. (2016). English medium instruction in

    universities: A collaborative experiment in Turkey. *Studies in English Language*

    Teaching, 4(1), 51-76.

Macaro, E. (2017). *English medium instruction: Global views and countries in focus*. Plenary

    speech presented at the symposium held at the Department of Education, University

    of Oxford.

Mackay, R. (1993). Embarrassment and hygiene in the classroom. *ELT Journal, 47*(1),

    32–39.

MacSwan, J., Thompson, M. S., Rolstad, K., McAlister, K., & Lobo, G. (2017). Three

    Theories of the Effects of Language Education Programs: An Empirical Evaluation of

    Bilingual and English-Only Policies. *Annual Review of Applied Linguistics, 37*, 218-

    240.

Madhavan, D., & McDonald, J. (2014) Webinar: English as medium of instruction (EMI): Philosophies and policies. *OECD Higher Education Programme*. Retrieved from: http://bit.ly/2gitABC

Marsh, D. (2002*). CLIL/EMILE – The European dimension: Actions, trends and foresight potential*. Finland, Jyväskylä: European Communities.

Marsh, D., & Wolff, D. (Eds.). (2007). *Diverse contexts – converging goals. CLIL in Europe.* Frankfurt: Peter Lang.

Manh, L. (2012). English as a medium of instruction in Asian universities: The case of Vietnam. *Language Education in Asia, 3*(2), 263-267.

Mannon, T. M. (1986). *Teacher talk: A comparison of a teacher's speech to native and non-native speakers* (Unpublished master's thesis). University of California, Los Angeles.

Mauranen, A. (2012). Second-order language contact: English as an academic lingua franca. In Filppula, M., J. Klemola, & D. Sharma (Eds.) *The Oxford handbook of world Englishes.* Oxford*:* Oxford University Press.

McMullen, M. G. (2014). The value and attributes of an effective preparatory English program: perceptions of Saudi university students. *English Language Teaching*, *7*(7), 131-140.

McField, G. P., & McField, D. R. (2014). The consistent outcome of bilingual education programs: A meta-analysis of meta-analyses. In G. P. McField (Ed.), *The miseducation of English learners* (pp. 267–298). Charlotte, NC: Information Age.

Mehisto, P., Marsh, D., & Frigols, M.J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.

Meisel, J. (1980). Linguistic Simplification. In S. Felix (Ed.). *Second Language Development: Trends and Issues* (pp. 13-40). Tübingen, Germany: Gunter Narr,

Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz

    & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual*

    education (pp. 35–63). Clevedon: Multilingual Matters.

Milk, R. (1982). Language use in bilingual classrooms: Two case studies. In M. Hines & W.

    Rutherford (Eds.), *On TESOL '81* (pp.181-191).Washington, DC.:TESOL

Mizon, S. (1981). *Teacher talk: A case study from the Bangalore/Madras communicational*

    ELT project. (Unpublished master's thesis). University of Lancaster, England.

Morell, T., Alesón, M., Bell, D., Escabias, P., Palazón, M., Martínez-Espinosa, R.M. (2014).

    English as the medium of instruction: A response to internationalization. Retrieved

    from: http://bit.ly/2f65sjH

Morris, S. (1983). Dictation: A technique in need of reappraisal. *ELT Journal, 37*(2), 121-

    126.

Nation, I. S. O. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.

Navés, T. & Victori, M. (2010). CLIL in Catalonia: An overview of research studies. In D.

    Lasagabaster, & Y. Ruiz de Zarobe (Eds.) *CLIL in Spain: Implementation, results and*

    teacher training (pp. 30-54). Newcastle: Cambridge Scholars Publishing.

Norris, J. & Ortega, L. (2009). Measurement for understanding: An organic approach to

    investigating complexity, accuracy, and fluency in SLA. *Applied Linguistics, 30*: 555–

    78.

Oh, S. (2001). Two types of input modification and EFL reading comprehension:

    Simplification versus elaboration. *Spring, 35*(1), 69–96.

Oller, W. J. (1971). Dictation as a device for testing foreign-language proficiency. *English*

    Language Teaching, 25(3), 254-259.

Oller, W. J. (1979). *Language tests at school: A pragmatic approach*. London: Longman.

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.

Ortega L (2012). Interlanguage complexity: A construct in search of theoretical renewal. In: B. Kortmann & B. Szmrecsanyi (Eds). *Linguistic complexity: Second language acquisition*, *indigenization, contact* (pp. 127–55). Berlin: de Gruyter.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*, 590–601.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117-134.

Papaja, K. (2014). *Focus on CLIL: A Qualitative evaluation of content and language integrated learning (CLIL) in Polish secondary education*. Newcastle: Cambridge Scholars Publishing.

Parker, K., & Chaudron, C. (1987). The effects of linguistic simplification and elaborative modifications on L2 comprehension. *University of Hawai'I Working Papers in English as a Second Language, 6,* 107-133.

Phillipson, R. (2009). *Linguistic Imperialism Continued*. London: Routledge.

Quené H., & van de Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425.

R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*(2), 45-105.

Robinson, P. (2005). Cognitive complexity and task sequencing: studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43*, 1-32.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics*, *45*(3), 161-176.

Rodríguez, J. (2012). Content and Language Integrated Learning (CLIL): Considerations in the Colombian context. *Gist Education and Learning Research Journal, 6*, 177-189.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*, 572–594.

Ruiz de Zarobe, Y., & Lasagabaster, D. **(**2010**)**. Introduction. The emergence of CLIL in Spain. An educational challenge. In D. Lasagabaster and Y. Ruiz de Zarobe (Eds.). *CLIL in Spain: Implementation, results and teacher training* (pp. ix - xvii*).* Newcastle*:* Cambridge Scholars.

Sabbour, S. M., Dewedar, S. A., & Kandil, S. K. (2010). Language barriers in medical education and attitudes towards Arabization of medicine: Student and staff perspectives. *Eastern Mediterranean Health Journal (EMHJ), 16*(12), 1263-1271.

Sawyer, J. O., & Silver, S. K. (1972). Dictation in language learning. In H. B. Allen & R. N. Campbell (Eds.), *Teaching English as a second language* (pp. 223-229). San Francisco: McGrawHill.

Schachter, J. (1996). Maturation and universal grammar. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 159–193). San Diego, CA: Academic Press.

Schierloh, M., & Paulsell, P. (2010). Input modifications by instructors in traditional and business German courses. *Global Business Languages*, *15*, 49-66.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206–226.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York, NY: Cambridge University Press.

Seliger, H.W. (1977). Does practice make perfect?: A study of interaction patterns and L2 competence. *Language Learning, 27*(2), 263-278.

Shamim, F. (2011). English as the language for development in Pakistan: Issues, challenges and possible solutions. In H. Coleman (Ed.), *Dreams and realities: English in development* (pp. 297-315). London: The British Council.

Shamim, F., Abduelhalim, A., & Hamid, N. (2016). English medium instruction in the transition year: Case from KSA. *Arab World English Journal (AWEJ), 7*(1), 32-47.

Shohamy, E. (2013). A critical perspective on the use of English as medium of instruction at universities. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English as medium of instruction at universities: Global challenges* (pp. 196-210). Bristol: Multilingual Matters.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428.

Slavin, R., & Cheung, A. (2005). A synthesis of research of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247–284.

Slobin, D. (1971*). Psycholinguistics* (2nd ed.). Glenview, Illinois: Scott Foresman & Co.

Smit, U., & Dafouz, E. (2012). Integrating content and language in higher education. An introduction to English-medium policies, conceptual issues and research practices across Europe. *AILA Review, 25*, 1-12.

Snow. C. (1972). Mothers' speech to children learning language. *Child Development, 43*, 549-565.

Södergård, M. (2008). Teacher strategies for second language production in immersion kindergarten in Finland. In T. Fortune & D. Tedick (Eds.), *Pathways to bilingualism and multilingualism: Evolving perspectives on immersion education* (pp. 152–73). Clevedon: Multilingual Matters.

Suliman, I. E. (2004). Arabicization in higher education: The case of medical colleges in the Sudan (Doctoral dissertation). Retrieved from University Microfilms International (UMI). (No. U601193)

Suliman, W., & Tadros, A. (2011). Nursing students coping with English as a foreign language medium of instruction. *Nurse Education Today, 31*, 402–407.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education.* Clevedon, England: Multilingual Matters.

Swain, M., & Lapkin, S. (1986). Immersion French in secondary schools: 'The goods' and 'the bads'. *Contact*, *5*(3), 2-9.

Tarnopolsky, O. (2013). Content-based instruction, CLIL, and immersion in teaching ESP at tertiary schools in non-English-speaking countries. *Journal of ELT and Applied Linguistics (JELTAL), 1*(1), 1-11.

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition, 12*(3), 287-301.

VanPatten, B. (1993). Grammar teaching for the acquisition-rich classroom. *Foreign Language Annals*, *26*(4), 435-450.

Vu, N.T., & Burns, A. (2014). English as a medium of instruction: Challenges for Vietnamese tertiary lecturers. *The Journal of Asia TEFL, 11*(3), 1-31.

Werther, C., Denver, L., Jensen, C., & Mees, I.M. (2014). Using English as medium of instruction at university level in Denmark: The lecturer's perspective. *Journal of Multilingual and Multicultural Development, 35*(5), 443-462.

Wesche, M.B., & Ready, D. (1985). Foreigner talk in the university classroom. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 89-114). Rowley, MA.: Newbury House.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Report No. 17). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Yeh, C. (2012). Instructors' perspectives on English-medium instruction in Taiwanese universities. *Curriculum and Teaching Quarterly, 16*(1), 209-232.

Yano, Y, Long, M., & Ross, S. (1994). Effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning, 44,* 189–219.

Zare-ee, A., & Gholami, K. (2013). Academic justifications for preferring English as a medium of instruction by Iranian university teachers. In M. Basri, K. Abd Rahman, M. Sofuan, M. Salleh, & N. Salleh (Eds.) *Proceeding of the Global Summit on Education (GSE2013)* (pp. 426-431). Kuala Lumpur: WorldConferences.