8-2019

Article Use of English Learners Classified by the Givenness Hierarchy

Elizabeth Claessens

Article Use of English Learners Classified by the Givenness Hierarchy

by

Elizabeth Claessens

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfilment of the Requirements

for the Degree of

Master of Arts

in English: Teaching English as a Second Language

August, 2019

Thesis Committee:
Edward Sadrai, Chairperson
Choonkyong Kim
Tim Fountaine

Abstract

The following research was conducted to understand the use of English articles in referential noun phrases (RNPs) by students from Nepal. English articles, “a/an” and “the”, are widely known to be one of the most difficult aspects of English to learn for second language learners of English. This thesis examines the article use of language learners by using the Givenness Hierarchy Framework, an implicational hierarchy of cognitive statuses proposed by Gundel, Hedberg, and Zacharski (1993). The data used for this study were written samples of students from Nepal, who had newly entered a university in the Midwest, who took a placement test at the beginning of their academic career at that University. Thirty student placement essays were the materials utilized for this research. The essays were transcribed, the referential noun phrases (RNPs) were identified and numbered, the article use in each RNP was evaluated for English-likeness, and the cognitive status of each RNP was recognized according to the Givenness Hierarchy. The results show the dispersion of the RNPs across the cognitive statuses within the Givenness Hierarchy. Additionally, the results show a high level of English-likeness in each cognitive status category, and frequent non-English-like variations of oversuppliance of “the”, deletion of “a/an”, and deletion of “the”. This research analyzes the cognitive status and English-likeness of the article use of students from Nepal.

Table of Contents

List of Figures

Chapter I: Introduction

Linguists may seem to disagree about nearly every area of the field, yet they rally around the difficulty of English articles for language learners, teachers, and researchers. Even after language learners labor in extensive study, live in a primarily English-speaking country for decades, and spend countless hours trying to remedy their difficulty with article errors, some errors remain. One of the reasons for this phenomenon is that a speaker can convey meaning in English without correct article use. When a language feature does not greatly inhibit communication, that language feature tends to remain an error in learner language. The use of English articles persists as a great difficulty for most English language learners.

According to Ekiert (2016), some linguists have paid special attention to the article errors of language learners who have no articles in their L1 (first language). Findings have included the frequent omission of articles by L2 (second language) learners. Ekiert (2016) reports that the omission of articles by L2 learners with no article in their L1 does not significantly improve with higher proficiency. English article errors are known to be a difficult obstacle to overcome for English language learners, especially for those with no articles in their L1.

According to Simons & Fennig (2018), Nepali, the principal language of Nepal, does not have articles. Nepali (sometimes called Nepalese) is the national language of Nepal, an Indo-Aryan language, which is located in South Asia between India and China. A 2001 census estimates world-wide Nepali L1 speakers at 17,209,255, and 11,053,255 people within the country (Summer Institute of Linguistics, 2005). Preliminary searches have not yielded any research focusing on the English article use of Nepali speakers. However, there are language learner studies focusing on other languages, including: Saudi Arabian Arabic speakers, Pakistani

Urdu speakers, Spanish speakers, and Chinese speakers (Al-Qadi, 2017; Ishaq & Bukhari, 2016; Dean, 2017). There is a need for research on Nepali L1, English L2 learner English article use.

The corpus that was used for this study was composed of written tests of international students enrolled at a midwest university. The type of meta-data available for this study can only identify the students' country of origin, rather than their L1. Approximately two thirds of Nepali people consider an Indo-Aryan language their L1 (The Editors of Encyclopaedia Britannica, 2018). The written tests of international students from Nepal will be analyzed in the following study.

Nepali is a lingua franca of Nepal, which is located in South Asia between India and China. Nepali is spoken throughout Nepal, with concentration in the West and the South (Summer Institute of Linguistics, 2005). The language class is: "Indo-European, Indo-Iranian, Indo-Aryan, Northern zone, Eastern Pahari" (Summer Institute of Linguistics, 2005, p. 276), with at least seven dialects. Speakers of Nepali include Hindus, Buddists, Muslims, and Christians. Nepali speakers also range four classes, from the highest class to the untouchable class. The Nepali language, spoken by millions in South Asia, is an Indo-Aryan language.

English articles (*a/an* and *the*), which are a difficult aspect of language to use for most English language learners (including L1 Nepali, L2 English learners), are frequently used referential forms. As humans communicate, they constantly refer to objects using referential forms such as *this*, *the*, and *a/an*; these referential forms are associated with accessibility and cognitive statuses. As individuals use language, they consistently use the "appropriate form to refer to a particular object" (Gundel, Hedberg, and Zacharski, 1993, p. 274). Both the context and the referential form used give enough information to the hearer for correct identification of

the “intended referent” (Gundel et al., 1993, p. 274). The intended referent determines the referential form, which can be described using the concept of accessibility (Ariel, 1991)

The object being referred to is accessible to the hearer, but the accessibility of each intended referent varies. Speakers refer to “mental representations” (Ariel, 1991, p. 433) which exist in short-term memory, long-term memory, or are newly constructed during the conversation. These mental representations could include a person, specific object, or idea. Though these mental representations are used frequently throughout conversation, the listener does not have “[equal accessibility]” (1991, p. 443) to them all. For example, the accessibility of a mutually known city between the speaker and listener is not as easily accessibile as a foreign city that is only known to the speaker. The accessibility of mental representations informs the use of referential forms, such as *the* and *a/an* (1991). Referential forms point to something, such as a person, and the referential form guides the listener in understanding the intended referent, by indicating the accessibility of that person, object, or idea.

The referential form, accessibility of the intended referent, and the cognitive status of the intended referent are all connected. When a speaker uses a referential form, he/she assumes the cognitive status of the referent within the hearer’s mind. Speakers refer to objects, they guide the hearer connect the words being spoken with the intended referent; referential forms (*a, the, it*) are used to help the hearer recall which object is being referenced. Referential language is used prevalently, which can only be used for successful communication with correct assumption of the accessibility and cognitive status for each referential expression.

According to Gundel et al. (1993), the referential form is associated with the cognitive status of the intended referent. Since “the distribution and understanding” (Gundel et al., 1993, p. 274) of referring expressions has been a historic dilemma for linguists, the Givenness Hierarchy

was proposed. According to these researchers, “different determiners and pronominal forms conventionally signal different cognitive statuses…thereby enabling the addressee to restrict the set of possible referents” (Gundel et al., 1993, p. 275). There is a clear relationship between referential forms (i.e. *the, this, a/an*) and the cognitive status of the intended referents. The Givennesss Hierarchy proposes a set of cognitive statuses, which are connected to referential forms.

Six cognitive statuses are recognized by the Givenness Hierarchy, which is an implicational hierarchy. These statuses are as follows: In Focus, the “most restrictive” (Gundel et al., 1993, p. 276), Activated, Familiar, Uniquely Identifiable, Referential, and Type Identifiable, the “least restrictive”. It is implicational in that every referential phrase is the least restrictive cognitive status (Type Identifiable). If the cognitive status of a referential phrase is recognized as Referential, it is also, by definition, Type Identifiable. Likewise, a phrase identified as Activated is also, by definition, Familiar, Uniquely Identifiable, Referential, and Type Identifiable. In this way, the hierarchy is implicational. This framework will be utilized in the following research to categorize the referential noun phrases (RNPs) within the corpus examined.

The data for this research project was writing samples written by thirty students of a university in the Midwest. The students had previously taken a written placement test at the time of their entrance into their university, to have their English proficiency evaluated. The students were all from Nepal, though their L1 can not be known with certainty. A grammaticality judgement evaluator evaluated the English-liked nature of the student’s article use within RNPs. The evaluator was an American, Minnesota resident, whose L1 was English, and who was both a trained writing tutor and a student of the graduate level Teach English as Second Language (TESL) program at the students’ university. The evaluator drew on both her native language,

local dialect, and professional expertise to make grammaticality judgements. The evaluator selected both the English-likeness for each RNP and the English-like variation for the non-English-like RNPs. Writing samples from thirty university students from Nepal and the evaluations of one L1 English evaluator were the ultilized for this research.

The materials utilized for this study were thirty written placement tests, which were prepared by the researcher. The placement test was given at the beginning of the students' academic career at their university, to evaluate whether or not they needed additional English language learning support via English for Academic Purposes (EAP) classes. All of the essays from the first week of testing of one semester, which were written by students from Nepal were selected. Thirty of those essays were randomly selected, transcribed by the researcher into a digital format, and the RNPs in each essay were underlined and numbered. These thirty essays, estimating a total corpus of 10,000 words were utilized for this research.

The thirty transcribed placement essays were delivered to the evaluator, who evaluated for English-likeness by the evaluator, and evaluated for each RNP's cognitive status by the researcher. Each essay included numbered RNPs, and numbered evaluative form. The evaluator read each essay, marking whether the article use in each RNP was English-like or non-English-like. For the RNPs that were marked non-English-like, the evaluator chose the well formed variation: "Insert/Replace with 'a/an'", Insert/Replace with 'the'", or "Remove article provided by the writer". The cognitive status of each RNP was recognized by the researcher, who used specific criteria to identify the cognitive status as defined by the Givenness Hierarchy. Both the English-likeness of each RNP, and the cognitive status of each RNP were identified for this research project.

The results of this research revealed the distribution of the RNPs across the Givenness Hierarchy, the percentage of English-likeness within each cognitive status, and the most frequent non-English-like variations of English article use. The total number of RNPs within each cognitive status exhibited the most prevalent cognitive status of RNP was Familiar, with 30% of the total, and the least prevalent cognitive status was Type Identifiable, with 9%. The English-likeness in each cognitive status remained above 90%, with the lowest percentage being Referential, with 91%, and the highest percentage being In Focus, with 98%. The most frequent non-English-like article use variations were, by great majority, the oversuppliance of "the", the deletion of "a/an", and the deletion of "the". The results for the RNP analysis for the corpus examined demonstrated a high proportion of Familiar RNPs, a overall high level of English-likeness, and the frequent oversuppliance of "the" and deletion of "a/an" and "the".

This report describes the process of seeking to better understand the article use of English language learners from Nepal. Chapter 2, the literature review, gives a history of analysis of English learner language, a history of classification of English articles, previous efforts to classify L2 English article errors, an overview of accessibility and cognitive status, an explaination of the Givenness Hierarchy, and the research question. Chapter 3, the methods section, describes the materials, the preparation and coding of materials, and the processes of both identifying RNPs and each RNP's cognitive status. Chapter 4, the results section, provides both written and visual representations of the distribution of RNPs in each cognitive status on the Givenness Hierarchy, the percent of English-likeness within each cognitive status, and the most prevalent non-English-like variations found within the corpus examined. Chapter 5, the discussion, restates the main idea of this research, summarizes the results, and relates this research to other areas of study. Chapter 6, the conclusion, explains the results, states the known

limitations to the study, and gives suggestions for further research. The following research report seeks to understand the article use of English langauge learners, as categorized by the Givenness Hierarchy.

Chapter II: Literature Review

The following literature review considers a brief history and background of the analysis of learner errors, analysis of English article errors, categorization of English article errors, accessibility, cognitive status, and the Givenness Hierarchy. Previous analyses of English articles have proven insufficient for understanding learners accurate use of articles. To understand articles and their use, accessibility and cognitive status must be considered. The Givenness Hierarchy may prove useful in analyzing the use of articles.

Analyzing Learner Language: Contrastive Analysis Hypothesis, Error Analysis, and Interlanguage

The analysis of the errors made by language learners has changed the way linguists view language development. The Contrastive Analysis Hypothesis was formed as linguists sought to accurately predict language learner errors based on the difference between each learner's L1 and L2. The Error Analysis framework was developed to study all learner errors, rather than just those that could be traced back to the learner's L1. As linguists began to understand how the produced language of learners is a complex system, they introduced the Interlanguage Hypothesis. These hypotheses have resulted from linguists' view of language errors, and guide linguists and teachers' understanding of language learner errors.

Proponents of "Contrastive Analysis Hypothesis (CAH)" (Lightbown & Spada, 2013, p. 41) set out to predict the errors of L2 learners based on the "transfer" from each learner's L1. CAH researchers compared each learner's L1 and L2, predicting learner difficulties based on the differences between the two (or more) languages spoken by the learner. Under the CAH, linguists, teachers, and learners assumed that the learner's L1 was, in part, responsible for language learner errors. For example, a major language in Nepal, Nepali, has no articles (Simons

& Fennig, 2018). So, if Nepali-speaking learners of English often failed to include English articles, it would be considered L1 transfer. CAH adherants identified the predicted "areas of difficulty" (Corder, 1967, p. 162) as they studied learners' L1 and the learners' target language. Target language (TL) is "the language being learned" (Lightbown & Spada, 2013, p. 223) which could be the learner's L1, L2, or an additional language being learned. Linguists' analysis of learners' L1 and TL could guide the language teacher's curriculum, as they seek to correctly predict areas of difficulty.

There are, however, shortcomings of the CAH framework. Teachers familiar with L2 learner errors found that many of the students' errors were not predicted by CAH (Corder, 1967). Rather, a vast number of errors could be attributed to "learners' developing knowledge of the structure of the target language" (Lightbown & Spada, 2013, p. 42). These errors associated with the structure of the TL could be found across learners from different linguistic backgrounds, rather than only from learners from the same linguistic background. While CAH was helpful in guiding researchers and teachers in the understanding of errors due to L1 transfer, errors originating from TL could not be explained by CAH. Furthermore, studies of L2 learners "imply that contrastive analysis may be most predictive" (Richards, 1970, p. 2) for the pronunciation of L2 learners. However, CAH is "least predictive" (Richards, 1970, p. 2) for L2 learner word order (i.e. syntax). The CAH framework is helpful for understanding L1 transfer, but it fails to address the errors that are related to the TL.

A new way of studying the errors of L2 learners, which later was named "error analysis" (Lightbown & Spada, 2013, p. 42), emerged due to CAH's failure to provide an explanation for all learner errors. Researchers working within the CAH approached L2 learner errors as evidence of L1 transfer. However, those adopting the error analysis framework discovered many language

learning errors which could be found in the language production of any language learner (Richards, 1970). Error Analysis sought to understand language learner errors which stem from both the L1 and the TL.

Language errors that are common in the produced language of any language learner are called "intralingual and developmental errors," (Richards, 1970, p. 2-3). These errors are not only found among L2 learners of the same L1 but also among L2 learners of disparate L1s, "deaf children learning" (Richards, 1970, p. 3-4) their written L1, and children learning their L1. The following errors are considered developmental: "*did he comed, what you are doing, he coming from Israel*, [and] *make him to do it*" (Richards, 1970, p. 3). These kinds of errors are seen across language learning, without direct connection to the learner's linguistic background. Error analysis explicitly focused on the errors produced by learners, rather than setting out to predict their errors (Lightbown & Spada, 2013).

The language produced by language learners is a "linguistic system" (Lakshmanan & Selinker, 2001, p. 395), which is "at least partially different" from the learner's L1 and the TL; this is known as the Interlanguage Hypothesis. As research of L2 errors was conducted, errors were found to be manifestations of the language learners' current knowledge of the TL's rules (Lightbown & Spada, 2013). Interlanguage is not fully their L1 nor the perfectly formed TL (Lightbown & Spada, 2013). Rather than viewing learner language as riddled with inaccuracies or viewing it as sub-par to the TL, interlanguage is viewed as its own "highly structured" (Selinker, 1969, p. 71) linguistic system. Interlanguage can be influenced by each learner's L1, L2(s) other than the TL, the TL, and developmental errors. Understanding the interlanguage hypothesis is crucial to grasping the analysis of learner errors since interlanguage has significant backing in the linguistic community of researchers (Lakshmanan & Selinker, 2001).

Linguists' approach to the analysis of language learner errors has changed over time, from seeking to predict errors to recognizing learner language as a complex linguistic system. Proponents of CAH sought to predict learner errors by analyzing the linguistic differences between the L1 and the TL. Those using the Error Analysis framework took all learner errors into account, including those resulting from developmental errors. In conjunction with the Error Analysis framework, the Interlanguage hypothesis was formed as linguists began to see learner errors as part of a complex linguistic system. The approach to the analysis of language learner errors has changed over time based on the gaps in hypotheses.

Classification Systems for English Articles

As mentioned earlier, understanding L2 learners' use of articles is pertinent for researchers due to the difficulty they pose for learners. The correct use of English articles is a significant obstacle for non-native speakers of English, even after many years of learning in an English speaking community. English articles are function words, which rarely obscure the meaning being communicated. This may be one of the factors that keeps article errors prevalent in L2 English learner language. Multiple researchers have proposed methods for explaining the correct use of articles, based on multiple binary decisions; two of these are outlined briefly below. Binary decisions for determining articles are easy to understand, which could make them easy for language learners to apply to their writing. However, these frameworks have not shown to be sufficient for English learners, considering that article errors persist.

There are several known causes for the difficulty L2 English learners have with articles. One of the factors that contributes to this difficulty is the lack of articles in some learners' L1 (Zhao & Macwhinney, 2018). Additionally, articles present difficulty in learning due to the fact that they seldom significantly obscure the intended meaning of the speaker (Master, 1990).

Language errors that do not inhibit the interlocutor from understanding the meaning tend to remain in L2 learner language, which prolong those language errors (Master, 1990). According to Master (1990), another reason for article difficulty is the "unstressed" (p. 461) nature of English articles. Even though there is a great need for teaching article usage, English language teachers often have no better explanation than whether it "sounds" (Master, 1990, p. 461) right or wrong. Therefore, linguists studying English have sought to find a satisfactory method of classifying English articles.

Master (1990) published a binary way of teaching articles to L2 learners. In Master's system, the learner needs only to make three decisions: "[±identified]" (1990, p. 470), "[±singular]", and "[±count]". Rather than the previous classifications of "[±definite] and [±specific]" (Master, 1990, p. 7), both of these classifications are combined into one with [±identified]. The word 'tick' in the phrase, "a tick carries disease" (Master, 1990, p. 7) is -identified, while, 'computer' in the prhase, "the computer is down today" is +identified. In this system, two classifications are brought together for the sake of simplicity in understanding English articles (Master, 1990). This system of classifying articles is considered more simple because it reduced the amount of decisions made by the L2 learner. By combining two "features" (Master, 1990, p. 470) into one, the number of decisions made by the L2 learner is diminished. This system is more simplistic than many of the other article classifications systems proposed, which could be helpful for language learners and teachers. In his system, the definite article: "the" is only needed if the noun is [+identified]. A noun which is [-identified], [+count], and [+singular] requires the indefinite article, "a". Finally, nouns which are [-identified] and either [+count] and [-singular] or [-count] have no article, Ø. As shown below, L2 learners need to make a maximum of three decisions to determine the correct article.

Decision tree for Determining the English Articles

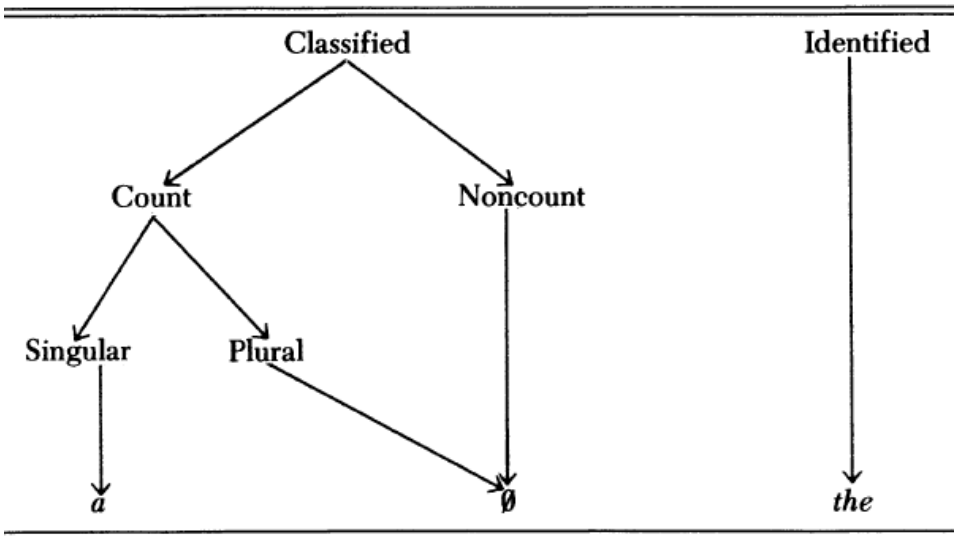

Figure 1: Diagram for Determining the English Articles (Master, 1990, p. 470)

Master (1990) provided a classification system for articles, which could make students' determining of English articles more straight-forward. However, this system does not account for all article uses in English, and learners taught this system continue to make frequent article errors (2018). According to Zhao and Macwhinney (2018), "pedagogical treatments of article instruction have not articulated an effective method for teaching" (2018, p. 100) English articles, as many highly proficient English learners "continue to produce errors in article usage" in their produced language. Article errors persist in advanced English learners, even with the above classification system.

A well known grammar book for English teachers by Celce-Murcia & Larsen-Freeman (1999) also uses a few noun features to help learners use articles correctly. The features used are: "common" (p. 272) or "proper", "count" or "noncount", and "singular" or "plural". Their system has similarities to Master's (1990), but Celce-Murcia & Larsen-Freeman (1999) made a distinction between common and proper, rather than between identified and classified. These

researchers give helpful frameworks for beginning to understand English article, but more recent researchers give more nuanced classification systems.

Systems for choosing which article to use have been developed to guide L2 English learners. These systems have had the goal of helping language learners by giving them a series of binary decisions, which could lead to the correct article choice. However, these systems have not solved the issue of prevalent article errors produced L2 learners. The lack of sufficient classification system for English articles led Gundel et al. (1993) to propose the Givenness Hierarchy. This classification system does not provide English learners a series of binary decisions. However, it can be used across languages to classify the NPs in which articles are used. The Givenness Hierarchy is further explored below.

Classifying L2 English Article Errors

Researchers have classified L2 English learners' use of articles in various ways. One way of classifying article use has been through accuracy only. Also, article use has been classified by the type of error, by considering the comparison between the correct article use and the provided article use. Finally, the type of NP based on definiteness and number has been used in conjunction with the type of error based on the correct usage. Several methods have been used to classify L2 English article errors, which are explained further below.

As researchers have sought to classify learners' article errors, they have used systems with a limited number of classification choices. Al-Qadi (2017) analyzed the language of L2 English speakers whose L1 was Arabic. These L2 English learners' article use was organized using three categories: correct, incorrect, and no attempt. There are only three categories used to study article use in this study. While this study can show the linguist the prevalence of article errors, it gives no specifications concerning the type of errors.

Ishaq & Bukhari (2016) also studied L2 English speakers' article errors, giving slightly more information about article use. The L1 of the participants was Urdu. Four categories were used for article errors: omission of "a", omission of "the", substitution of "a/the", and substitution of "Ø/the" (Ishaq & Bukhari, 2016). This study gives the linguist some insight into the kind of article errors that may occur. However, these categories do not give information about the context of the error, since there are many instances where "a" could be omitted, for example. These researchers classified the type of article error by the difference between the produced article and the correct article.

Dean (2017) classified article errors by English language learners using a two-tiered system which includes definiteness. Article errors are categorized by: "singular definite" (Dean, 2017, p. 76), "plural definite", "singular indefinite". In Dean's (2017) work analyzing English article use, she uses *definite* to refer to *the* and *indefinite* to refer to *a/an*. She does not give further definition of these terms within this research. The NPs in the corpus examined are further classified as "obligatory context" (Dean, 2017, p. 76), "correct suppliance", and "oversuppliance". While this does give more specific information than the former studies, there is still want of a more specific classification system for English NPs.

While these studies provide some understanding of how prevalent article errors are, there is little information concerning which kind of article errors are most commonly correct or incorrect. Rather, the accuracy alone is considered. The article errors compared to the correct use is analyzed. However, there is a need to have an understanding of article errors that can guide English teachers' treatment of articles within the classroom.

Accessibility and Cognitive Status

As humans communicate, they constantly refer (or point) to people or things that vary in accessibility and, therefore, cognitive status (location in memory) of the referent. As individuals refer to something or someone, the listener must decipher which something or someone is being referred to. Not all objects have the same degree of accessibility in the hearer's mind. The speaker often uses referents, such as *the*, to signal the hearer as he/she recalls or forms a representation of the referent NP. The accessibility is directly related to the cognitive status of the hearer, as it is assumed by the speaker. Accessibility and cognitive status are essential to understanding and processing NPs.

As individuals use language to communicate, they use words (and body language in spoken discourse) which are interpreted by the listener through the access of "mental representations", according to Ariel (1991, p.443). During human communication, mental representations that are in short-term memory, long-term memory, and or newly created are accessed in the course of conversation. Not all mental representations are "equally accessible" (Ariel, 1991, p. 443) to the listener during a discourse. For example, both of the mental representations of one's high school teacher and one's current employer are accessible to a forty-year-old. However, the latter is certainly more readily accessible than the former. These "mental representations (specifically those of NPs) are accessible to [listeners] in varying degrees" (Ariel, 1991, p. 443). As speakers use NPs to refer to objects, the hearer's accessibility to the NPs vary at the point at which the NP is uttered by the speaker.

The accessibility of an NP is partially signaled by the "referring expressions" (Ariel, 1991, p. 443) such as *this*, *him*, and *the*. Referring expressions point to an object, which could be a person, group of people, thing, place, etc. The listener is signaled to the "degree of

accessibility" (Ariel, 1991, p. 444) by referring expressions. Referring expressions guide the listener as he/she recalls or forms the object being referred to.

Language speakers use referents, such as *this*, *that*, *the*, *a* and *it* within NPs to refer to something; these referents are directly associated with cognitive statuses (Gundel et al., 1993). "What do speakers/writers know that enables them to choose an appropriate form to refer to a particular object?" (Gundel et al., 1993, p. 274). The answer to this question is: speakers use language in a way that gives the hearer all the information needed to "identify correctly the intended referent"(Gundel et al., 1993, p. 278). Language users use referents frequently, which could refer to almost anything, yet they rarely need to inquire: "what were you referring to?" According to Gundel et al. (1993), these referents are connected to specific cognitive statuses. "Cognitive status is a measure of what the speaker may assume about the listener's knowledge, beliefs, and attention state in the context where the referent is mentioned" (Swierzbin, 2004, p. 5). Referents can be frequently and accurately used throughout language because each referent is directly related to the cognitive status of both the speaker and the listener.

A basic understanding of accessibility and cognitive status gives context for the cognitive statuses which are crucial for understanding the Givenness Hierarchy. As speakers use referential language to point to an object, the hearer is able to correctly associate the particular object with the referential expression. Referential NPs vary in accessibility to the hearer, as he/she forms or recalls the correct object. The degree of accessibility is directly connected to the cognitive status of each object (in the hearer's mind) that is referred to, as assumed by the speaker. The referents used within referential NPs are tied to the cognitive status of the NP.

Givenness Hierarchy

Linguists have historically had difficulty with "the distribution and understanding" of different forms of referring expression" (Gundel et al., 1993, p. 274). This difficulty lead Gundel et al. (1993) to propose the Givenness Hierarchy framework. As they state, "different determiners and pronominal forms conventionally signal different cognitive statuses…thereby enabling the addressee to restrict the set of possible referents" (Gundel et al., 1993, p. 275). The referent used (i.e. *this*, *the*) in any given context is dependent on the "assumed cognitive status of the referent" (Gundel et al., 1993, p. 275); this relationship between the cognitive status and the referent form is "widely recognized" by linguists. The Givenness Hierarchy recognizes six cognitive statuses, which range from "most restrictive (in focus) to least restrictive (type identifiable)" (Gundel et al., 1993, p. 276). The following diagram, taken directly from this research , outlines the six proposed cognitive statuses and the relevant forms associated with each status.

(1) THE GIVENNESS HIERARCHY:

in focus	> activated	> familiar	> uniquely identifiable	> referential	> type identifiable
{*it*}	{*that*, *this*, *this* N}	{*that* N}	{*the* N}	{indefinite *this* N}	{*a* N}

Figure 2: The Givenness Hierarchy (Gundel et al., 1993, p. 275)

A cognitive status is the assumed location of a referent in the addressee's memory. When a speaker/writer uses a referent, such as *this* or *the*, he/she "assumes the associated cognitive status is met" (Gundel et al., 1993, p. 275). For example, two strangers could be speaking, and use the phrase "the moon" without a previous mention of this in their conversation. Though the speakers are not known to each other, each is able to assume that this new acquaintance knows

exactly which object is being referred to by the NP “the moon”. As referents are used, the cognitive status of the referent is assumed in the hearer/reader’s mind.

Within the Givenness Hierarchy there are six cognitive statuses, each of which have defined relationships to each other. When a speaker/writer uses a referent, which is always associated with a cognitive status, “all lower statuses (to the right in the Hierarchy) are met as well” (Swierzbin, 2004, p. 26). Therefore, In Focus referents must also be Activated, Familiar, Uniquely Identifiable, Referential, and Type Identifiable (Gundel et al., 1993). When a referent such as “the moon” is Familiar, it must also be Uniquely Identifiable, Referential, and Type Identifiable. The relationships between the statuses are explained below.

The definitions of each cognitive status help linguists better understand the nature each of the cognitive statuses, but differentiating the statuses from each other is complex. Labeling referential NPs with a specific cognitive status can be difficult due to their closely-related nature. The following explanation of each of the statuses includes guidlines of how to determine a cognitive status for and NP. The statuses within the Giveness Hierarchy can be used to label referential NPs with specific guidelines.

1. “In Focus” (Gundel et al., 1993, p. 275) is the highest level, or most restrictive, cognitive status in the Givenness Hierarchy. In Focus referents are both in “short term memory” (Gundel et al., 1993, p. 279) and “at the current center of attention”. The hearer/reader does not need to recall which object is being referenced due to its place in the conversation. Rather, it is currently at the forefront of the hearer/reader’s mind. Therefore, the referent’s cognitive status is In Focus.

An In Focus referent can be determined if:

In Focus	"the referent was mentioned in a syntactically prominent position (e.g., main clause subject) in the immediately preceding sentence" (Sadrai, 2016, based on personal communications with Jeanett Gundel).
	"the referent was a higher-level topic that was part of the interpretation of the immediately preceding sentence" (Sadrai, 2016).
	"the referent was mentioned earlier in the same sentence" (Sadrai, 2016).

Example 1 from Gundel et al. (1993):

"a. My neighbor's bull mastiff bit a girl on a bike.

b. It's the same dog that bit Mary Ben last summer" (Gundel et al., 1993, p. 280).

In Example 1 sentence b, the NP "It" is In Focus because when the speaker uses the referent "it", the specific dog is at the "current center of attention" (Gundel et al., 1993, p. 279). A linguist analyzing the above sentence can know that the referent is In Focus because it was mentioned in the sentence right before the sentence containing the In Focus referent. The object being referred to by the referent "it" is quite clear to the listener/reader.

2. The next cognitive status, Activated, is in short term memory, or is in the "extralinguistic context" (Swierzbin, 2004, p. 29). The object which is being referred to could be in the physical area of the speakers; or, the referent could be mentioned recently in the discourse. However, Activated referents are not at the center of attention, which differentiates them from In Focus referents. Activated referents are in immediate environment, either linguistically or physically, or, they are in short term memory.

An Activated referent can be determined if:

Activated	"the referent was mentioned in the previous two sentences" (Sadrai, 2016).
	"the referent was present in the immediate, non-linguistic spatio-temporal (i.e., physical) context" (Sadrai , 2016).
	"the referent was mentioned three sentences previously, but not necessarily since then" (Sadrai, 2016).

Example 2:

"My neighbor has a dog. That dog kept me awake last night" (Gundel et al., 1993, p. 279).

In Example 2, the specific dog was mentioned in the immediately preceding sentence and, therefore, in short term memory. The NP "That dog" is Activated, as demonstrated by the recent mention of the referent. However, the mention of the specific dog in the first sentence is not in a syntactically prominent position. "That dog" is mentioned in the immediately preceding sentence, but not in a syntactically prominent position. Hence, this referential NP is Activated but not In Focus.

3. Familiar referents can be specifically identified by the hearer/reader due to an already-existing "representation" (Gundel et al., 1993, p. 278) in his/her mind. The Familiar referent could be an already-existing specific object in the hearer/listener's mind. Additionally, Familiar referents can be from common knowledge or experiences. Thirdly, this type of referent could have been mentioned in the previous discourse; in the case of a written work, two references to the same object could be pages or chapters displaced from one another. While this referent already can be specifically identified by the audience, the speaker/writer does not assume that this referent is in the hearer/reader's short-term memory. When a

speaker/writer uses a referent classified as Familiar, he/she assumes that the addressee has an already-existing specific object in his/her mind.

A Familiar referent can be determined if:

Familiar	"the referent was mentioned at any time previously in the discourse" (Sadrai, 2016)
	"the referent was assumed to be shared knowledge between speaker and hearer (either through shared cultural knowledge or shared experience)" (Sadrai, 2016)

Example 3:

"I couldn't sleep last night. That kept me awake" (Gundel et al., 1993, p. 278).

The NP "That" in Example 1, which is assumed to be shared knowledge between the speaker/writer and hearer/reader, is Familiar. The speaker's use of "that" in this context would only be "appropriate" (Gundel et al., 1993, p. 278) if the hearer/reader The speaker assumes that the listener has previous knowledge of the object which is being referred with "That". In this context, the speaker would be either referring to something in the immediate extra-linguistic context, or has recently mentioned the object. By nature of the Givenness Hierarchy, "That" is Familiar, Activated, and In Focus. The speaker of Examle 3 would only use the Familiar referent "that" in the context of those two sentences if he/she assumed that the hearer/reader knew of this neighbor dog.

4. The Uniquely Identifiable status is met if the referent is known either through previous knowledge, or if there is sufficient new information to "associate a new unique representation with the referent" (Sadrai, 2016). These referents can be pointed out by the hearer specifically, based only on the "nominal" (Gundel et al., 1993, p. 277). The addressee's correct identification of the referent that is intended by the speaker/writer can be "based on

an already existing representation" (Gundel et al., 1993, p. 277) within his/her memory. The addressee can identify the object mentioned either by previous knowledge, or by sufficient new information to compose a new unique representation in the hearer's mind. The addressee, however, does not need the rest of the sentence to associate this new representation within his/her memory. A Uniquely Identifiable referent can be correctly pointed out by the addressee as he/she associates a specific representation in the memory.

A Uniquely Identifiable referent can be determined if:

Uniquely Identifiable	"the referring form contained adequate descriptive content to construct a new unique representation" (Sadrai, 2016)
	"the hearer was able to identify a unique referent by linking it indirectly to a recently activated referent" (Sadrai, 2016)

Example 4:

"I couldn't sleep last night. The dog (next door) kept me awake" (Gundel et al., 1993, p. 277).

In Example 4, the NP "the dog" is Uniquely Identifiable. "The dog" has not been previously mentioned in the discourse, and the hearer does not need any previous knowledge of a neighbor dog to understand uniquely identify the object being referred to. However, the speaker gives enough context within the "nominal alone" (Gundel et al., 1993, p. 277) for the hearer to construct a new representation. "The dog" is not Familiar, and the rest of the sentence is not needed to compose a Uniquely Identifiable representation of the dog. The addressee can construct a new and unique image of a Uniquely Identifiable referent from the nominal used by the speaker/writer.

5. The Referential cognitive status is met when the speaker is referring to a "particular entity" (Swierzbin, 2004, p. 28). Though the object being referred to is not already known to the hearer, he/she can only construct a new representation after the sentence has been processed, and the hearer must "access an appropriate type-representation" (Gundel et al., 1993, p. 276). The Referential cognitive status is met when the speaker refers to a particular object, unlike Type Identifiable referents. Additionally, this status is met when the hearer needs to process the entire sentence to construct a new represetation in his/her memory, unlike the Uniquely Identifiable status, in which the hearer needs only the nominal to construct a new representation. When the speaker refers to a particular object which is not already known by the hearer, and the hearer needs the full sentence to "I couldn't sleep last night. This construct a new representation of this object, the Referential cognitive status is met.

A Referential referent can be determined if:

Referential	"the referent was mentioned subsequently in the discourse" (Sadrai, 2016)
	"no subsequent references to the entity occur, but it is reasonably clear from the context that the speaker intends to refer to a particular object" (Sadrai, 2016)

Example 5:

"I couldn't sleep last night. This dog (next door) kept me awake" (1993, p. 277).

In Example 5, the NP "This dog" meets the Referential cognitive status. The hearer would already know what a dog is; additionally, the hearer would have a new representation of this dog when the speaker has completed the sentence. In Example 5, the speaker is not necessarily assuming that the hearer knows of the particular dog prior to this interaction;

however, the speaker gives the addressee enough information to construct a new representation before the end of the sentence.

6. Type Identifiable referents give the hearer enough information to identify the type of object being referred to by the speaker/writer. The addressee needs only the knowledge of the kind of object which is being mentioned by the speaker/writer. This is the cognitive status necessary for any NP to be considered referential; therefore, all referential NPs meet this cognitive status.

A Type Identifiable referent can be determined if:

Type Identifiable	"the hearer can identify the type of thing described, i.e., the hearer can be assumed to understand the meaning of the words and the sense of the expression in general" (Sadrai, 2016).

Example 6:

"I couldn't sleep last night. A dog (next door) kept me awake" (Gundel et al., 1993, p. 276).

In the above example, "A dog" is Type Identifiable. The dog is not necessarily singular, and the speaker/writer is not referring to a specific dog. While the speaker/writer is referring to a dog, neither the speaker/writer nor the addressee necessarily knows which dog kept the speaker/writer from sleeping. However, the hearer can identify the type of object mentioned. The indefinite article used in this NP signals to the addresee that some dog(s) was keeping the speaker/writer awake, but it may have been one of several dogs. The only information that the hearer/reader needs is to know what "dog" means in general.

The Givenness Hierarchy identifies six cognitive statuses, which have been used to identify referential expressions across several languages (Gundel et al., 1993; Swierzbin, 2004).

They can be utilized to better understand the article use of English language learners, as cognitive statuses are connected to the articles used. The most restrictive cognitive status, In Focus, is met when the object being referred to is both in the short-term memory of the hearer and at the current center of attention. The least restrictive status, Type Identifiable, is met when the speaker is only expected to identify the type of object being referred to; the speaker does not expect the hearer to have a particular object in mind. Every referential use of language is necessarily Type Identifiable, but may or may not be any more restrictive than that cognitive status. As cognitive statuses move from Type Identifiable to be more restrictive, "all lower statuses (to the right in the Hierarchy) are met as well" (Swierzbin, 2004, p. 26). Therefore, a Familiar referent is also, necessarily, Uniquely Identifiable, Referential, and Type Identifiable. However, Familiar referents are not Activated or In Focus. The Givenness Hierarchy provides a framework for classifying all referential NPs, which correspond to the referent form used by the speaker, such as *this*, *the*, and *a/an*. These cognitive statuses has be utilized to analyze produced language in multiple languages; the Givenness Hierarchy includes all referential uses of language.

The Givenness Hierarchy will be employed to analyze the corpus examined; the specific aspect of language to be analyzed is English articles. This framework will be helpful in the analysis of English article use because it will provide a picture of how English learners use referential language across the six cognitive statuses. English article use will be analyzed using the Givenness Hierarchy.

Preliminary searches have found that Swierzbin's (2004) study is the only research to use the Givenness Hierarchy to specifically analyze English article use; the following study will contribute to that area of research. Swierzbin's (2004) research analyzes the referential NPs used

by English learners, which includes a broader analyses of not only English articles, but also determiners, personal pronouns, quantifiers, and other referential forms. In her analysis, she analyzed the language of Japanese speakers who spoke English as an L2. Swierzbin (2004) used the Givenness Hierarchy to analyze L2 learners of English, including analysis of L2 learners article use and accuracy.

English articles have historically been quite difficult for L2 English learners. The analysis of learner errors, as well as the analysis of article errors is examined. There are numerous ways linguists have classified English articles, however, their frameworks fall short of successfully guiding learners' article use. Articles are a referential form, which correspond directly to the accessibility and cognitive status of the hearer. The Givenness Hierarchy is helpful for categorizing L2 learner referential NPs along a taxonomy. The article use of learners of English can be analyzed using the Givenness Hierarchy.

Research Question

How do Nepali students, who are newly entering an American university use English articles, as classified by the Givenness Hierarchy?

Chapter III: Method

Participants

The writing of students from Nepal who were newly entering an American University was used for this research; one evaluator participant was utilized to mark the English-likeness of each referential noun phrase (RNP).

Second Language Learner Participants. University students who had met their University's proficiency requirements took a written placement test, which was utilized for this research. While the students had met the University's proficiency requirements for admittance, they were tested immediately before their first semester of their University experience in the United States. The test was utilized to decide whether or not students needed additional English for Academic Purposes (EAP) classes in addition to their university program requirements. These essays were used for the following research, and permission for access to these essays was specifically requested. Written proficiency tests from L2 language learners made up the corpus examined in the following study.

The participants were students enrolled at a Midwest University who took a placement writing test immediately before beginning their first semester of university in the United States. The group was made up of thirty students, whose combined writing samples were expected to total approximately 10,000, based on an expectation of 300 words per student essay. The students writing totaled 11,241, with a mean of 374.7 words per essay. The students were made up of both males and females, all of whom were from Nepal. The corpus used for this research was composed of the written placement tests of students from Nepal.

All of the participants met the University's English proficiency requirements prior to taking this written test; however, the results of this test would determine whether or not an

additional EAP class was needed for each student. The University's proficiency requirements could be met by several methods of measuring proficiency, including a score of 61 on the TOEFL (iBT), completing the University's Intensive English program, or graduating from an English medium high school with a minimum GPA of 2.5. The purpose of this test was to decide if the students needed EAP reading and writing instruction, which they would take concurrently with their major's coursework requirements. The outcome of this test impacted their choice of courses, but did not inhibit them from taking the courses needed for their program's requirements.

The written proficiency test, given by the university's English department, was securely archived after the test was given, scored, and recorded. After the test was completed by the students, they were given to the EAP Graduate Teaching Assistants (GTAs) for scoring. The scores of the students were recorded by the GTAs, and then the tests were given to the EAP Coordinator. The purpose of the tests was met at this point, but the tests were kept by the EAP Coordinator for any future reference or research. With the tests in storage, individuals could request access for research through the Institutional Review Board (IRB).

To obtain access to these exams, the researcher requested permission from the IRB of the university and the EAP Program Director. The students' written exams were all taken from one test, given at the beginning of one semester, with the same prompt. The semester was chosen based on the number of Nepali students, a semester with at least thirty students from Nepal was be chosen. That semester's test record had fourty-nine students from Nepal, who had taken the test during the first week of testing, and whose written tests were immiately found by the EAP Coordinator when searching through the stored exams. Those fourty-nine essays were numbered 1-49, and then a random number generator (Random.org) was used to select thirty essays

randomly. Those thirty, randomly selected, written exams were used for this study, to the exclusion of all writing samples from non-Nepali students. For the use of these students' writing samples for this research, the IRB and the EAP Director gave permission.

The L2 learners whose language was analyzed for this study were L2 English learners from Nepal. These L2 learners met their University's proficiency requirements and were enrolled at a Midwest university. The students took a written proficiency test, to evaluate their potential need for additional English learning support during their academic career. These written tests were utilized for the following research.

Native-English-Speaking, Writing Specialist Evaluator. A grammaticality judgement evaluator, who was an L1 English-speaker, was asked to evaluate the English-like vs. non-English-like nature of the articles used throughout the corpus examined. The evaluator was an American L1-English speaker who had been trained to work in an on-campus writing center, and was in the MA Teaching English as a Second Language (TESL) graduate program.These qualifications provided both native-knowledge of English and professional training for consistency for the evaluator's decisions.

The evaluator lived in Minnesota and was American; she also is a native speaker of English. This geographic specification of living in Minnesota was made due to the dialectical variation within native English speakers. The University of the L2 learners is located in Minnesota. There is bound to be dialectical variation among university faculty, students, and Minnesotan variations of English. However, the dialect used at the University can be categorized as generally Minnesotan and academic. This English speaker knew English from childhood. Finding a completely monolingual individual with the above qualifications was not obtainable. The requirement of being a native English speaker was made to guide the English language

professional through both her skills and her language background. The evaluator used both her intuition of the Minnesotan dialect of the English language to make these decisions, as well as her professional English language training.

Finally, the grammaticality judgement evaluator was trained to work in the on-campus writing center and was in the MA TESL graduate program a the university. Since the data being evaluated was written for the academic setting of an American university, Standard American English was expected. The students who wrote the EAP placement essays were enrolled in an American university; the essays were written for the purpose of assessing the students' academic writing. Therefore, the writing evaluated was expected to be academic and Standard American English by the program giving the assessment. The grammaticality judgement evaluator was an individual with training in academic English, due to the academic nature of the writing being evaluated.

A trained tutor from the writing center on-campus and MA TESL student was selected for the grammaticality judgement evaluator because she was trained in academic Standard American English. The writing center tutor was trained to guide students in their academic writing, but did not have specific training in article use. The evaluator was very familiar with the type of English expected in acadamia, and corrected articles based on whether or not the article use was considered appropriate in the academic setting. Additionally, as an MA TESL student, this individual was preparing to be an English teacher. Therefore, she also was familiar with, and preparing to teach, academic English. This evaluator was not given a specific set of article rules to follow, rather, she evaluated the data based on her intuition which was informed by her L1 of English and expertise in academic Standard American English. Based on these qualifications, this individual evaluated the English-likeness of L2 learners' English article use.

Materials

Student written tests. For the written placement tests, a prompt was written and proctored by University faculty. The prompt was passage-based, created by the director of the English for Academic Purposes (EAP) program at the University. The passage was a summary of an article on a subject matter that was accessible to all students, from the director's perspective (J. Condon, personal communication, October 19, 2018). The prompt was argumentative in genre, in that it instructed the students to form an opinion and write a five paragraph essay to argue that opinion. The facilitator explained the prompt to the students, emphasizing the question, time limit, and method of grading (J. Condon, personal communication, October 19, 2018). Next, the facilitator gave the students fifty minutes to complete the task, and they were notified when there were ten minutes and five minutes left. Students were given a printed passage, prompt, and lined writing paper. Before the students complete their essay, they were instructed to write their school identification number on their test, but not to put their name on the test. This task was hand-written; students had no access to outside resources. This placement test was created and proctored by the University's EAP program faculty.

Text preparation and coding of NPs. The corpus examined was prepared by the researcher, readied for evaluation by the grammaticality judgement evaluator. The identifying information of each student was removed, and the essays were reproduced in a digital format. The referential noun phrases (RNPs) in the corpus examined were identified by the researcher. The corpus was specifically prepared by the researcher for the confidentiality of the L2 learner participants and ease of use by the evaluator.

The written placement tests, after being obtained by the researcher, were prepared for evaluation. The identification numbers on the hand-written tests were removed, to protect the

students' identities. The tests were transcribed by the researcher, to convert the essays to a digital format. Next, the researcher carefully read each hand-written essay against the digital copy, checking to ensure the accuracy of the transcription. The corpus examined did not have any identifying information of the L2 learner participants, and was in digital form.

Next, the researcher finished preparing the corpus for evaluation by marking each RNP, and the data was evaluated by the grammaticality judgement evaluator. All of the RNPs within the data were identified; they were considered referential if they refered to something or someone (see the Givenness Hierarchy section of Chapter 2). Each RNP was underlined and numbered with superscript. A numbered table was created that corresponded to the numbered RNPs in the data. For each essay, a corresponding evaluation form was created and printed. The only individuals with access to the data were the grammaticality judgement evaluator and the researcher. The essays were printed with the respective evaluation form, and delivered to the evaluator. After evalutation, they were returned to the researcher for recording.

Written placement tests from the L2 learners were compiled into a corpus which was examined for this research. The students' identification numbers were removed from the essays, and the essays were transcribed by the researcher and checked against the original essays. The RNPs were prepared for the evaluator by identification, underlining, and numbering. Finally, the data and evaluative forms were printed and delivered for the data collection from the evaluator.

Procedures

After the corpus examined was prepared, the evaluator was given written directions (see Appendix B). These directions were to specify purpose of the evaluation. Next, the evaluator callibrated with the researcher to a minimum of 90% consistency.

Directions for RNP decisions were given to the grammaticality judgement evaluator in preparation for the evaluation of the corpus (see Appendix B). The researcher emphasized that the evaluator was only concerned with the English-like vs. non-English-like nature of the article use, and the English-like article variation. She was not concerned with any other part of the writing. The grammaticality judgement evaluator was given directions before evaluating the data for this research.

The evaluator was also prepared to evaluate with callibration with the researcher. The evaluator and researcher made decisions about the English-likeness of each RNP of the same essay. Both individuals decided whether each Referential Noun Phrase (RNP) in the corpus examined was English-like or non-English-like. Their answers were compared, with a requirement of 90% of the same decisions, for callibration.

After the evaluator was given directions and callibrated with the researcher, she was given the coded data for evaluation. The evaluator was permited to complete the data evaluation in multiple sessions, separated by days, to avoid unnecessary fatigue, which could potentially impact judgement. The essays were given to the evaluator in two portions, which were evaluated on two separate days. This allowance was given to prevent uneccesary fatigue in evaluating the data.

An evaluative form (see Appendix C) required the evaluator to determine whether the English article use in each RNP was English-like or non-English-like. If the article use was non-English-like, the evaluator could choose whether to insert/replace with “the”, “a/an”, or remove the article provided by the writer. For example, if the evaluator decided that RNP 36 (below) was non-English-like, then she decided whether to replace the student-written “the” with “a/an” or no article. Similarly for NP 37, if the evaluator decided that the article was non-English-like, she

would choose whether “the” or “a/an” should be inserted, where the student had not written an article. For each RNP, the evaluator chose whether each one is English-like or non-English-like, and if necessary chose the article English-like variation, and marked those choices using the evalutative form provided

For example:

So in [21]India [22]we are learn [23]the primary level of [24]English ... so [25]that's why... and then [26]we coming to [27]La Roche and ...[28]we know [29]the grammatical English so then [30]we don't need to do the from [31]beginner to[32] advanced level, know [33]little bit [34]English so [35]we can start from [36]the advanced level or [37]high beginner.

Evaluative Form:

21	English-like article	Non-English article		
		Insert/replace with “the”	Insert/replace with “a/an”	Remove article provided by writer
22	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
23	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
24	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
25	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
26	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
27	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
28	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer

29	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
30	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
31	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
32	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
33	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
34	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
35	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
36	English-like article	Non-English-like article		
		Insert/replace with "the"	Insert/replace with "a/an"	Remove article provided by writer
37	English-like article	Non-English-like article		
		Insert/replace with "the"	Insert/replace with "a/an"	Remove article provided by writer

The corpus examined was prepared for the grammaticality judgement evaluator and then given to her for evaluation. The evaluator made decisions as to whether each RNP was English-like or non-English-like, as well as the English-like variation of article use.

Analysis

Three primary forms of analysis for the purpose of this research were: the cognitive status of the corpus RNPs, the English-likeness of the article use within each RNP, and the type of non-English-like article variations generally occurring. All of the RNPs were identified and categorized based on the Givenness Hierarchy scale by the researcher. The cognitive status

signaled by each RNP was identified and labeled by specific criteria (see the Analysis section of Chapter 3 for the criteria). Additionally, each RNP was evaluated by the grammaticality judgement evaluator. Each RNP was evaluated as either English-like or non-English-like. Additionally, the grammaticality judgement evaluator gave the well formed choice of article use. The corpus was analyzed according to cognitive status, English-likeness, and well-formed article use.

Identifying Referential Noun Phrases. As RNPs were identified throughout the data, a few factors were considered. First, NPs containing expletive (dummy) pronouns were considered non-referential and excluded from this study. Second, noun phrases can contain verb phrases (VPs) and prepositional phrases (PPs), and as nouns occured within those phrases, the nouns within those VPs and PPs were identified seperately from the larger noun phrase, with some exceptions. Thirdly, when a language use error inhibited the researcher's ability to identify whether or not the phrase was referential, the phrase was excluded from the data. These principles guided the RNP decisions made throughout the data.

NPs containing expletive (dummy) pronouns, which tended to occur as subjects at the beginning of sentences, were recognized as non-referential. There were some instances of "there" or "it" at the beginning of sentence is usally not referring to a particular object, but is an expletive (otherwise known as dummy subject). In Essay 33 (below), there is both a referential, and a non-referential NP.

Essay 33, sentences from third paragraph:

"[61]Living off [62]campus is like a fish in a pond who thinks [63]that's [64]his whole world, but infact *there* is so much more out [65]there. *There* is so much to be explored and experienced." (italic emphasis added).

The first "there" in the above phrase is an expletive subject: "… infact *there* is so much more out [65]there". This first does not refer to a certain object or person, and is not referential. However, the second "there", is referential in that it points a place other than the campus. Expletive constructions, including some instances of "it" and "there" are not counted as RNPs throughout the data.

Due to the prevalence of nouns within different kinds of phrases, the identification of RNPs proved to be a more complex process than anticipated. For example, in the above exerpt from Essay 33, "[61]Living off [62]campus" is a RNP. Inside of this NP is a prepositional phrase: "off [62]campus"; within that prepesitional phrase is a noun: "[62]campus". Within many RNPs in the data, there were multiple phrases, including verb phrases and prepositional phrases. Within those verb phrases and prepositional phrases there were nouns. Decisions were made for each RNP taking consistency and maximum counts of nouns in mind.

As RNPs were identified, when multiple nouns were clearly being referred to by one referential form (such as *the, a, her, their,* etc.), those nouns were identified together within the same RNP. For example, the following RNP from Essay 30 (full sentence below) "[27]the different religions, cultures, and languages" was identified as one RNP (rather than three) because "different" is referring to the following three nouns.

Essay 30, sentence from second paragraph:

"[23]Living on [24]campus at [25]SCSU can help [26]a students to interact with [27]the different religions, cultures, and languages from all over [28]the world."

Contrastingly, in the following portion from Essay 31 (full sentence below), "[4]greater success and [5]engagement ", the student's exact intended meaning is somewhat ambiguous. The student may be referring to "greater success" and "greater…engagement"; however, the student

may be referring to “greater success” and “engagement” in general. These two examples of RNPs demonstrate how they were identified within the data.

Essay 31, first sentence:

“It is not uncommon to believe that [1]living on [2]campus in [3]university or college housing leads to [4]greater success and [5]engagement with [6]the university or [7]college community.”

Due to the nature of the prompt, there were certain phrases that were prevalent throughout the data, which required specific rules for identifying RNPs. “Campus housing” was considered one RNP. Two RNPs were identified the phrase “Living off campus”, while “Living off-campus” was considered one RNP. “Off” in the former example is a prepostion, while “off” in the latter example is part of a hyphenated word. These are examples of RNP identification decisions that were made throughout the data.

There was one recorded instance of a language use error in the data that inhibited the researcher’s ability to identify whether or not a phrase was referential. In Essay 23 (below), there were language use errors that resulted in abiguity.

Essay 23, sentence from the fourth paragraph:

“[82]They can involves [83]themselves *college community*, can have [84]greater acceptance for [85]diversity, and can increase [86]their personal growth and [87]strong connections to [88]campus life.”

The researcher was not able to identify whether or not the phrase “college community” was referential. In this instance, that phrase was thrown out of the data. When this language use error inhibited the meaning of this phrase, to the extent that it could not be identified as referential or not, that phrases were not used in the data. It is possible that there were more

instances of language use error inhibiting the meaning to the extent that the researcher could not know whether or not the phrase was referential; however, this was the only recorded instance of this situation.

The identification of each, individual RNP throughout the thirty student essays proved to be a more complex process than anticipated. As the researcher made decisions about each RNP, expletives were not included in the RNP count. Similarly, one phrase was not clearly referential or non-referential, due to a language use error, which was also not included in the RNP count. As noun phrases in English regularly include other phrases, such as VPs and PPs, the student's intended meaning was considered. Each RNP was identified according to these qualifications.

Identifying the cognitive status of each RNP. Gundel et al. (1993) proposed the Givenness Hierarchy, which allows linguists to categorize referring expressions in naturally-occurring language (readers are directed to the Givenness Hierarchy section of Chapter 2 above for an outline of this taxonomy.) This framework was utilized in this research to categorize the RNPs in the writing of students from Nepal. This hierarchy categorizes individual referents in language on a six-tier implicational scale of cognitive statuses. These range from In Focus (a referent that is at the forefront of the listener's mind) to Type Identifiable (identifying the type of object being described). The researcher categorized each RNP in the data on this six-tier scale.

For each cognitive status recognized in the Givenness Hierarchy, there are not only specific definitions, but also guidelines for determining the cognitive status for each RNP. These guidelines were used to consistently identify each RNP with a cognitive status. For each RNP, the researcher began with the guidelines for determining the Uniquely Identifiable cognitive status. If the cognitive status was not Uniquely Identifiable, the researcher moved to the more restrictive or less restrictive status, based on the RNP. The researcher continued to move farther

away from Uniquely Identifiable until the appropriate cognitive status was found. The guidelines were followed, to help ensure consistency in identification of the cognitive status of each RNP.

The cognitive status of each referent was identified using the categories recognized by the Givenness Hierarchy (Gundel et al., 1993). These statuses are defined in the Givenness Hierarchy section of Chapter 2, and each of the statuses was recognized by specific qualifications. The qualifications, as well as examples of each cognitive status are given below. The referents of certain cognitive statuses, such as In Focus and Activated, were more easily identified than the other cognitive statuses. The decisions made for the referents in each cognitive status, the straightforward identifications of cognitive statuses, as well as the more difficult identifications are outlined below.

Identifying In Focus RNPs:

In Focus	"the referent was mentioned in a syntactically prominent position (e.g., main clause subject) in the immediately preceding sentence" (Sadrai, 2016, based on personal communications with Jeanett Gundel).
	"the referent was a higher-level topic that was part of the interpretation of the immediately preceding sentence" (Sadrai, 2016).
	"the referent was mentioned earlier in the same sentence" (Sadrai, 2016).

In Focus referents, generally, were straightforward to identify, including the following example. In the below sentences from Essay 3, RNPs 6, 10, 11, and 12, were identified as In Focus.

Essay 3, first three sentences:

> "[1]Saint Cloud State University located at [2]720th 4th Ave is [3]a university of [4]opportunities. [5]Lots of 'International' and 'National' students are present at [6]SCSU. [7]First thing [8]a student must be prepared of is [9]the place where [10]they shall be staying as soon as [11]they reach [12]SCSU."

RNP 6, "SCSU" was "mentioned earlier in the same sentence" (Sadrai, 2016), as RNP 6 is the abbreviation used for RNP 1: "St. Cloud State University". RNPs 10 and 11 were identified as In Focus because that referent, "they" was mentioned in RNP 8, "a student", which occurred within the same sentence. RNP 12 was considered In Focus because the referent, "SCSU", was the subject in the previous sentence. RNP 12 was recognized as In Focus, since the referent "mentioned in a syntactically prominent position in the immediately preceding sentence" (Sadrai, 2016). This is an example of RNPs which were identified as In Focus. In the corpus examined, In Focus RNPs tended to have either no referential form, or a referent that was not an article, such as *their*, *those*, or *his*.

Identifying Activated RNPs:

Activated	"the referent was mentioned in the previous two sentences" (Sadrai, 2016).
	"the referent was present in the immediate, non-linguistic spatio-temporal (i.e., physical) context" (Sadrai, 2016).
	"the referent was mentioned three sentences previously, but not necessarily since then" (Sadrai, 2016).

Activated RNPs, similar to the In Focus status, are more straightforward to recognize than many of the other cognitive statuses. In Essay 3, RNP 18 and 19 are Activated.

Essay 3, sentences 2-4:

> [7]First thing [8]a student must be prepared of is [9]the place where [10]they shall be staying as soon as [11]they reach [12]SCSU. As its better to stay in [13]an organized dorm rather to spend [14] your night at [15]park until you get [16]a room. Yes, [17]I do think that [18]SCSU should require [19]students to live in [20]university housings but not for [21]the first year of [22]college, had it been for [23]a semester [24]students would have probably got familiar with [25]the college community."

If a referent was named in the two sentences preceding its use, then it is considered Activated. RNP 18, "SCSU", was also mentioned in RNP 12; RNP 19, "students", was mentioned in RNP 18. These two referents are mentioned in the previous two sentences; however, they are not In Focus. They are not In Focus because their previous mention is not in a "syntactically prominent position" (Sadrai, 2016) in the sentence previous. RNPs 18 and 19 are examples of the Activated cognitive status, according to the Givenness Hierarchy.

In the following exerpt from Essay 5, RNP 31 has been categorized as Activated.

Essay 5, sentence within second paragraph:

> If any of [31]the students face any [32]problem, [33]they can directly contact to [34]anyone but while living off [35]campus [36]they may have to call [37]911.

One of the ways Activated RNPs can be identified is when "the referent was mentioned in the previous two sentences" (Sadrai, 2016). RNP 31, "the students" was mentioned in the sentence immediately preceding this sentence. Therefore, this RNP can be categorized as Activated.

Identifying Familiar RNPs:

Familiar	"the referent was mentioned at any time previously in the discourse" (Sadrai, 2016)
	"the referent was assumed to be shared knowledge between speaker and hearer (either through shared cultural knowledge or shared experience)" (Sadrai, 2016)

Since the prompt for the essay was text-dependent, the students read a one-paragraph summary of a passage. The students wrote in response to the text and the corresponding questions. As the students wrote their responses, they wrote with the context given by the prompt. So, as Familiar RNPs were identified, those which were already mentioned in the prompt were identified as Familiar, even at their first mention in each student's essay. In Essay 28 (below), RNPs 18, 19, 21, 22, and 23 are considered Familiar.

Essay 28, second sentence:

> "Being one of [17]the well-established university in [18]Minnesota state [19]St. Cloud State University should require [20]students to live in [21]university housing during [22] their first year of [23]college."

The RNPs "Minnesota state", "St. Cloud State University", "university housing", "their first year" and "college" were considered Familiar, since they were mentioned previously in the prompt's text. Though the writer did not mention these referents previously, they were mentioned specifically in the prompt. The place of each referent on the Givenness Hierarchy is based on the writer's assumption of the referent in the mind of the readers. Therefore, the writer could safely assume that the reader knew the prompt which the essay was in response to. Therefore, the writer's use of referents including "university housing" and "college" are Familiar, due to their previous mention in the writing prompt.

In Essay 5 (below), RNP 37 is Familiar because the writer can assume that the reader has this shared experiential or cultural knowledge.

Essay 5, sentence within second paragraph:

> If any of [31]the students face any [32]problem, [33]they can directly contact to [34]anyone but while living off [35]campus [36]they may have to call [37]911.

RNP 37: "911" is the number called for emergencies in the US; this is considered common knowledge in the US. Therefore, "911" was considered "shared cultural knowledge or shared experience" (Sadrai, 2016). The writer, he/she wrote an essay within a US university, correctly assumed that the reader would know what "911" was. This is one example of how familiar RNPs were identified.

Examples of Familiar RNPs that were not very clearly identifiable were RNPs 31-35 in Essay 24.

Essay 24, exerpt from second paragraph:

> "Firstly, if [22]we stay [23]oncampus [24]we will be able to know about [25]the environment of [26]campus and [27]location of [28]different buildings. For instance, [29]we can easily know about where [30]different buildings lies (where is [31]our Garvey, [32]Newman centre, [33]Bradly Hall, [34]Department of Biology, [35]Department of Engineering and computing and [36]many more building) related to [37]our majors and minors."

The above sentences were located in the second paragraph of the essay, but the first mentions of "our Garvey", "Newman centre", "Bradly Hall", "Department of Biology", "Department of Engineering and computing". Each of these RNPs is a building or department on the campus on which the student studied, and the reader of the placement essay worked. The qualifications for Uniquely Identifiable RNPs could appear sufficient to recognize these first mentions of buildings

on the students' campus as Uniquely Identifiable; with the context, there might be enough "descriptive content to construct a new unique representation" (Sadrai, 2016). However, the student writer did not introduce these buildings as if they were new to the reader. Rather, the student mentioned them in passing, without any introduction. "our Garvey" and the other RNPs in the above phrase were categorizedas "shared cultural knowledge or shared experience" (Sadrai, 2016). Accordingly, RNPs 31-35 were recognized as Familiar, because of the shared knowledge between the reader and writer. Similar to In Focus examples of RNPs, there were not many examples of Familiar RNPs with any articles, they tended to have no referential form, as can be seen in the above examples.

Identifying Uniquely Identifiable RNPs:

Uniquely Identifiable	"the referring form contained adequate descriptive content to construct a new unique representation" (Sadrai, 2016)
	"the hearer was able to identify a unique referent by linking it indirectly to a recently activated referent" (Sadrai, 2016)

The first time each essay writer mentioned "I", "me", "we", or "us", those referents were considered Uniquely Identifiable. Though this decision is not very simple, it is the most appropriate choice on the Givenness Hierarchy in this situation. The writer can assume that the reader understands he/she is a student, writing the placement essay. The above pronouns were not mentioned previously in the text, nor are they shared knowledge between the reader and his/her audience, which are qualifications for Familiar RNPs. Though these pronouns could be considered shared knowledge, they better fit the description for Uniquely Identifiable. The reader has the ability to connect "I" to the writer, as the indended referent to which "I" is referring.

Pronouns such as “I”, “me”, “we”, and “us” are recognized as Uniquely Idenitifiable the first time they are mentioned in each essay.

In the following sentence from the data is a clear example of a Uniquely Identifiable RNP. In Essay 30 (below), RNP 27 is recognized as Uniquely Identifiable. With the details included within the RNP, as well as the immediately surrounding context of the sentence, the reader can “construct a new unique representation” (Sadrai, 2016). The student is referring to “different religions, cultures, and languages” on the university campus, which are “from all over the world”. That is sufficient information for the reader to “construct a new unique representation” (Sadrai, 2016).

Essay 30, sentence from second paragraph:

> “[23]Living on [24]campus at [25]SCSU can help [26]a students to interact with [27]the different religions, cultures, and languages from all over [28]the world.”

Identifying Referential RNPs:

Referential	“the referent was mentioned subsequently in the discourse” (Sadrai, 2016)
	“no subsequent references to the entity occur, but it is reasonably clear from the context that the speaker intends to refer to a particular object” (Sadrai, 2016)

In the following sentence from the data, RNPs 2 and 4 are recognized as clearly referential.

Essay 30, first sentence:

> “[1]A better accomodation in [2]a new place lead towards [3]the effective development for [4]the changed mind.”

Both RNPs 2 and 4 “refer to a particular object” (Sadrai, 2016). Also, the referents were not “mentioned subsequently” in this essay. These RNPs were not categorized Uniquely Identifiable

because the reader could not “construct a new unique representation” (Sadrai, 2016). Futhermore, RNPs 2 and 4 were not considered only Type Identifiable, because the reader can identify more than just “the type of thing described” (Sadrai, 2016). By this reasoning, RNPs 2 and 4 are considered Referential.

RNP 54 in the following sentence from Essay 5 is an example of a Referential RNP that was difficult to categorize.

Essay 5, exerpt from second paragraph:

> “In [53]my hometown, all of [54]the universities have made compulsory to [55]the freshman to live in [56]university housing.”

This is the only mention of “the universities” in the student’s hometown within the essay; therefore, the RNP is not Familiar. The writer mentions “the universities” in passing, introducing a new referent. The cognitive status of this RNP is difficult to identify, in part, because this is a new referent that the researcher might identify as Uniquely Identifiable, since the reader might be able to “construct a new unique representation” (Sadrai, 2016). However, as the student writer only mentioned this referent in passing, with little additional information, there is not enough information for the reader to construct a new representation. Hence, this RNP is not Uniquely Identifiable. Since “no subsequent references to the entity occur, but it is reasonably clear from the context that the speaker intends to refer to a particular object” (Sadrai, 2016), this RNP has been categorized as Referential.

Identifying Type Identifiable RNPs:

Type Identifiable	"the hearer can identify the type of thing described, i.e., the hearer can be assumed to understand the meaning of the words and the sense of the expression in general" (Sadrai, 2016).

In the following exerpt from the data, RNP 1 is clearly Type Identifiable.

Essay 30, first sentence:

"[1]A better accomodation in [2]a new place lead towards [3]the effective development for [4]the changed mind."

There is no mention in the writing prompt or later in this essay, of "a better accomodation". The reader "can identify the type of thing described" (Sadrai, 2016), which describes the Type Identifiable cognitive status. However, the reader is not refering to "a particular object" which confirms that this RNP is not Referential. RNP 1 in Essay 30 is an example of a straightforward identification of the Type Identifiable cognitive status on the Givenness Hierarchy.

In the following exerpt from the data set, RNPs 55 and 56 are Type Identifiable. These phrases are examples of cognitive status decisions that were not straightforward.

Essay 30, sentences in third paragraph:

"[47]The first year of [48]college can be [49]the best experience ever and [50]living on [51]campus at [52]SCSU can be at [53]its best. [54]learning requires [55]a peaceful environment and [56]an beautiful mindset and [57]SCSU is perfect in making [58]students to grab at [59]their fullest."

Neither before nor after this sentence are there any mentions of "a peaceful environment" or "a beautiful mindset" in the text. These phrases could easily be confused for Uniquely Identifiable RNPs because the reader might "construct a new unique representation" (Sadrai, 2016) based on

his/her own idea of such environments or mindsets, connected to a university campus. However, there is not enough information given at any point in the text for the reader to formulate a "new unique representation" (Sadrai, 2016). However, the student writing does not give any verbal signals that he/she is referring to a particular "peaceful environment". Rather, the student is referring to "a peaceful environment" in general. Since "the hearer can identify the type of thing described" (Sadrai, 2016), these RNPs were identified as Type Identifiable

Each of the six cognitive statuses recognized by the Givenness Hierarchy can be identified by the above criteria. While the statuses In Focus and Activated are relatively straightforward to identify, the statuses Uniquely Identifiable and Referential are more complex to identify. As decisions were made to identify the cognitive status of each RNP, those general decisions are outlined above. With the criteria given to categorize each RNP, they can be identified with consistency.

Results

The placement writing test was made up of a text-dependent prompt; the prompt was answered with a written response. These written responses, specifically of thirty students from Nepal, were used as the data for this research. All of the Referential Noun Phrases (RNP) were identified and then evaluated based on whether it was English-like or non-English-like. For the non-English-like phrases, the evaluator selected her variation of article choice. Next, the researcher went through the data, identifying the cognitive status of each RNP as explained by the Givenness Hierarchy.

The following visual representations (Figures 1, 2) give the percent of each cognitive status recognized by the Givenness Hierarchy. This is the primary answer to the research question: How do Nepali students, who are newly entering an American university use English

articles, as classified by the Givenness Hierarchy? The cognitive status of each Referential Noun Phrase within the corpus examined was identified. The following chart shows the proportion of each cognitive status within the corpus. Of the six cognitive statuses, Familiar is the most prevalent, with 30.09%. The least-represented cognitive status is Type Identifiable, with comprises 6% of the total. The percent of the RNP percentage of each cognitive status, according to the Givenness Hierarchy.

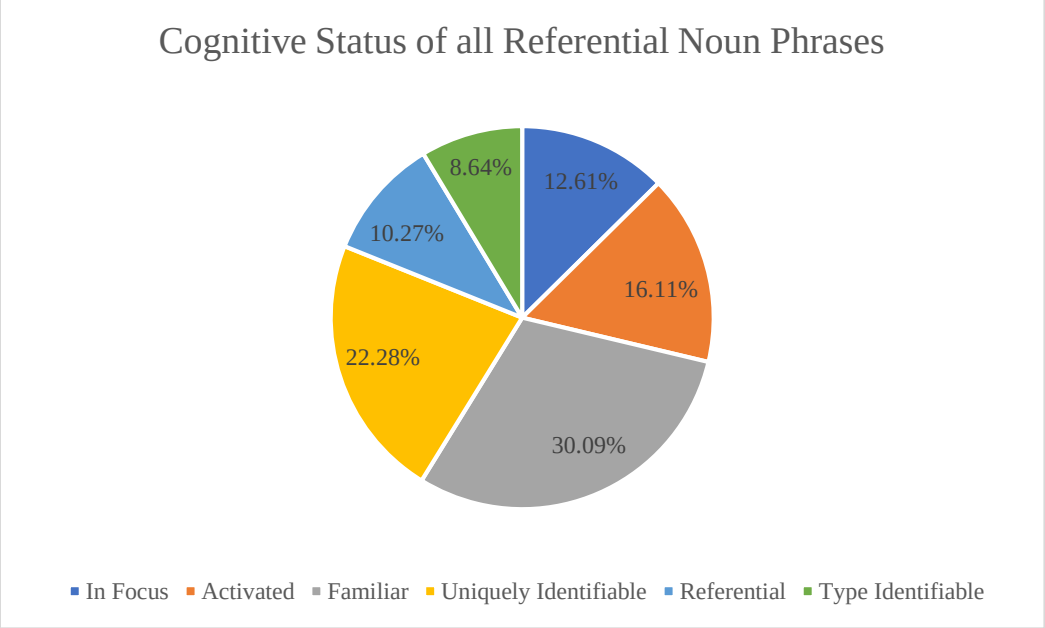

Figure 3: Distribution of all RNPs, shown in pie chart, across the Givenness Hierarchy

Cognitive Status	Percent of Whole
In Focus	12.61%
Activated	16.11%
Familiar	30.09%
Uniquely Identifiable	22.28%
Referential	10.27%
Type Identifiable	8.64%

Figure 4: Distribution of all RNPs, shown in table, across the Givenness Hierarchy

In the following two diagrams, all of the RNPs are separated by English-likeness or non-English-likeness. In Figure 5, the RNPs that were recognized as English-like by the grammaticality judgement evaluator. Due to the small number of non-English-like RNPs, compared with the English-like RNPs, Figure 5 has only slight differences compared to Figure 3. In Figure 6, the RNPs that were recognized as non-English-like by the evaluator are represented in a diagram. This chart does vary from Figure 3, in part, due to the small proportion of non-English-like RNPs. In total, there were only 160 non-English-like RNPs, compared to the total 2,998 RNPs. One of the noticable differences between these charts is the higher percentage of Referential RNPs in Figure 6. Referential was the cognitive status in which there was the highest percentage of non-English-like RNPs, as can be seen in Figure 7. This accounts for the difference in percentage of Referential RNPs between the two below figures.

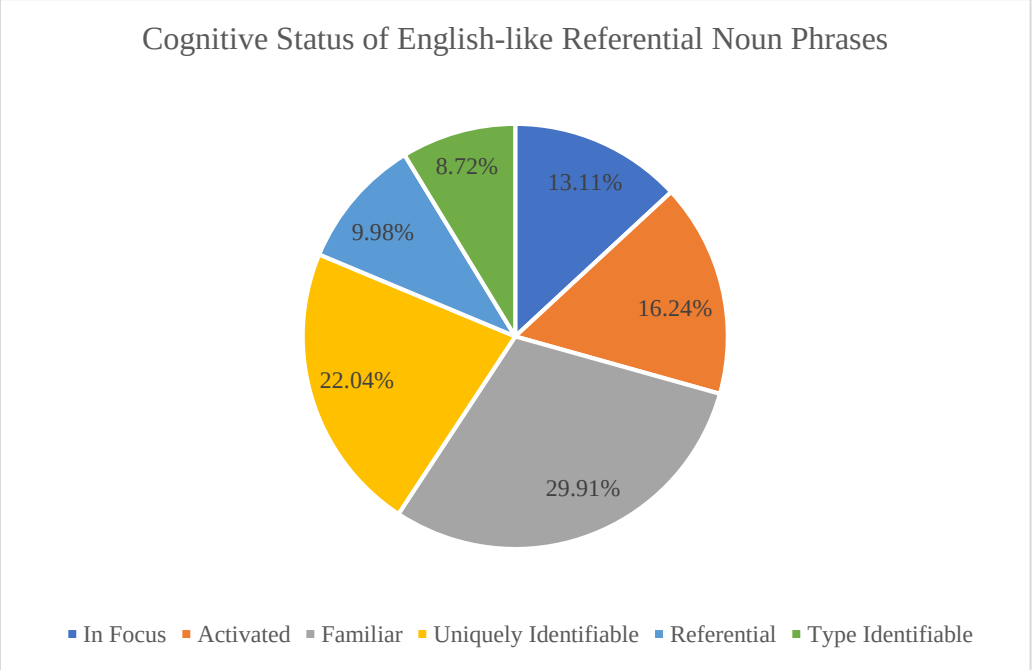

Figure 5: Distribution of English-like RNPs, shown in pie chart, across the Givenness Hierarchy

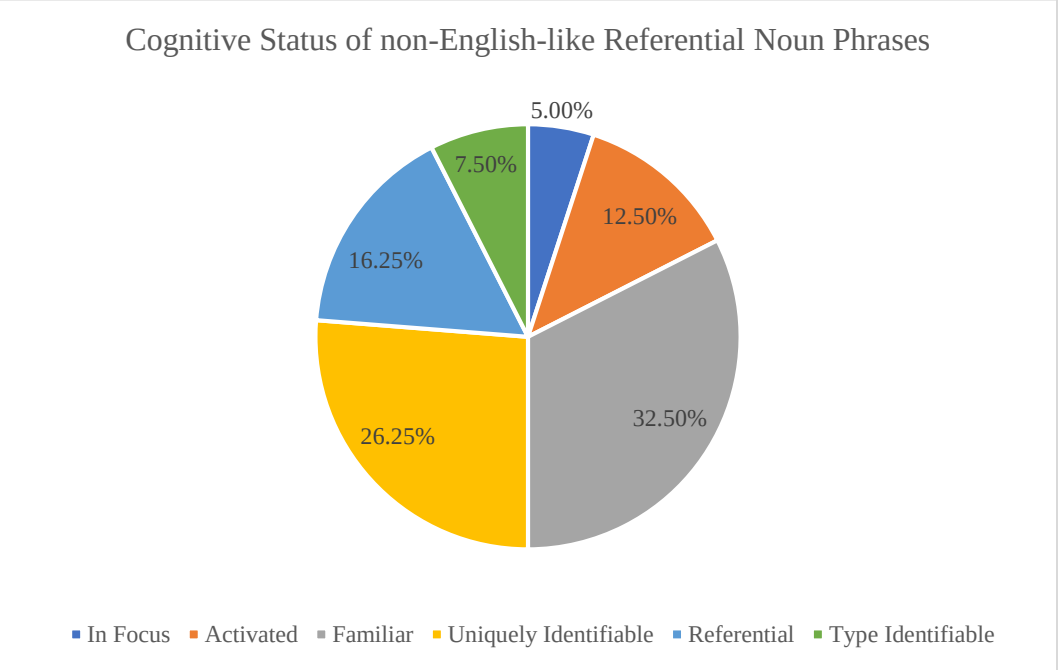

Figure 6: Distribution of non-English-like RNPs, shown in pie chart, across the Givenness Hierarchy

The bar graph in Figure 7 is a visual representation of the percent English-like within each cognitive status. While there is variance between the English-likeness within each cognitive status, there is notable consistency. This is seen in the range of the percent English-like: 98.14% for In Focus, compared to the 91.59% English-like for Referential. Each cognitive status' percent of English-likeness remains above 90%. This demonstrates a remarkable level of English-like use of English articles for this group of English Language Learners.

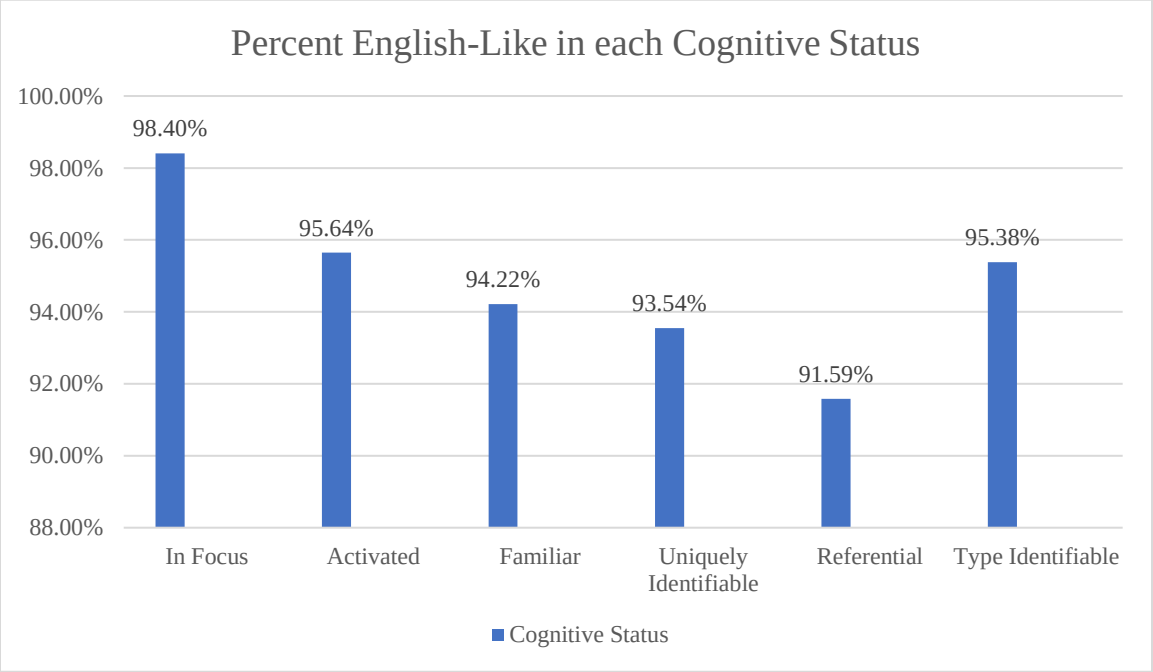

Figure 7: For each cognitive status, the percent English-like

As the evaluator read each student essay, she marked either English-like or non-English-like; for the non-English-like RNPs, she marked the English-like variation. For those non-English-like, the evaluator marked "Insert/replace with 'the'", "Insert/replace with 'a/an'", or "Remove article provided by writer". While there was not great variation between those three choices, the choice marked most often was "Remove article provided by writer". Figure 8 represents the percent of each choice made for the non-English-like RNPs.

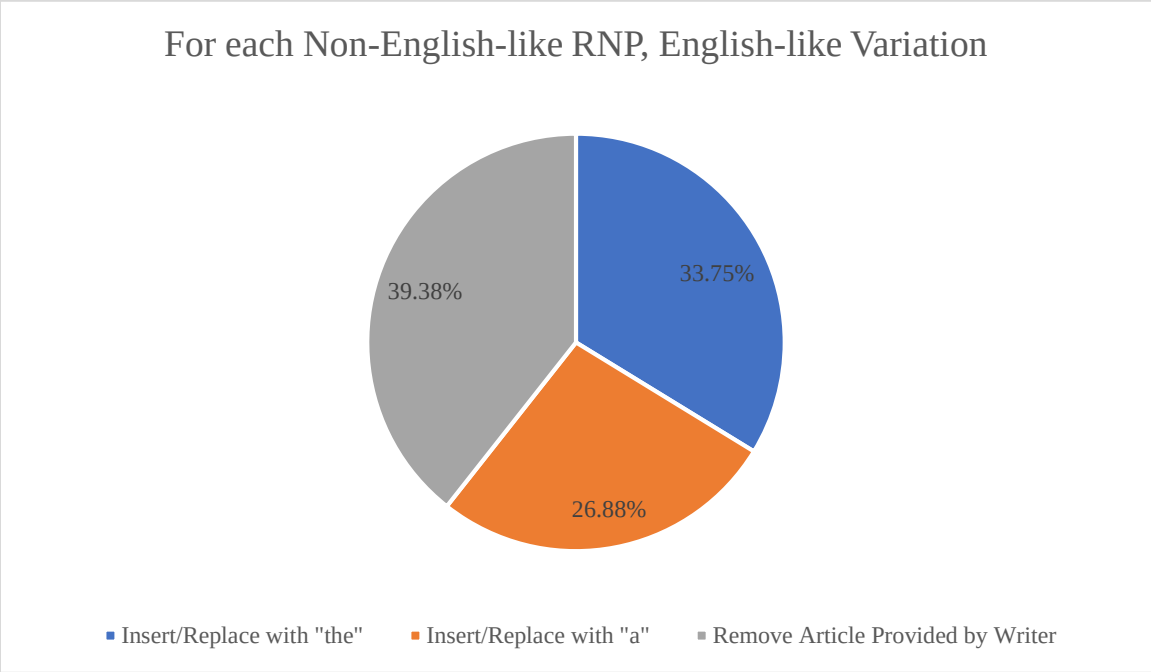

Figure 8: Non-English-like RNPs, distribution of English-like variation

Of all the non-English-like RNPs where the evaluator chose "Remove article provided by writer" as the English-like variation, the chart in Figure 9 shows the referent form given by the writer. In Essay 16 (below), RNP 47: "[47]the diversity" is an example of when the evaluator chose "non-English-like", and the student had written "the" before the noun. As seen in the chart below, of all the RNPs where the evaluator chose "non-English-like" and "Remove article provided by writer", 75% of those RNPs had a "the" written by the student. The 4.69% "Other", includes the instances such as RNP 71 in Essay 16 (below). This section includes referential forms such as "many researches", which were marked non-English-like. Of all the RNPs where the English-like variation was changed, by the evaluator, to "Remove article provided by writer", 72.73% had "the" written by the student. The oversuppliance of "the" was a significant non-English-like variation within the data examined.

Essay 16, sentence in second paragraph:

"Also, [37]speaking about [38]acceptance for [39]diversity, [40]students themselves while [41]attending campus, [42]participating in [43]events and [44]communicating and [45]interacting with [46]others will learn to accept [47]the diversity."

Essay 16, sentence in third paragraph:

"During [70]their studies, there are going to be [71]many researches to work on, [72]they will get [73]assignments to complete."

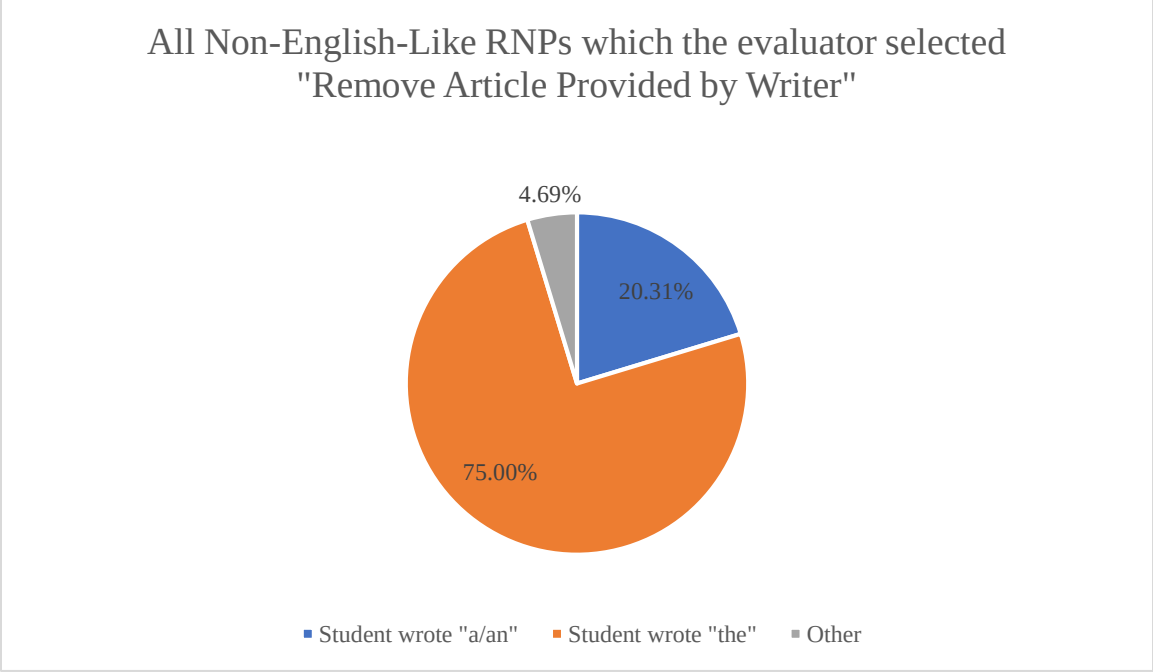

Figure 9: Non-English-like RNPs, for which no article would be included in the English-like variation

For each non-English-like RNP where the evaluator chose "Insert/Replace with 'a/an'" as the English-like variation, Figure 10 shows the original referent form given by the student. In Essay 16, RNP 66: "[66]student", is an example of when the evaluator chose "non-English-like",

and the writer had no article, referent form, or other word before the noun. As seen in the chart in Figure 8, of all the RNPs where the evaluator chose “non-English-like” and “Remove article provided by writer”, 90.24 % of those RNPs had no article or other word written by the student. Of all the RNPs where the English-like variation was changed, by the evaluator to "Insert/Replace with 'a/an'", 90.24% had no article or other word written by the student. The deletion of “a/an” was a notable non-English-like variation within the corpus examined.

Essay 16, sentence in third paragraph:

[64]These are [65]the things that will help [66]student in [67]their personal growth are [68]stronger connections to [69]campus life.

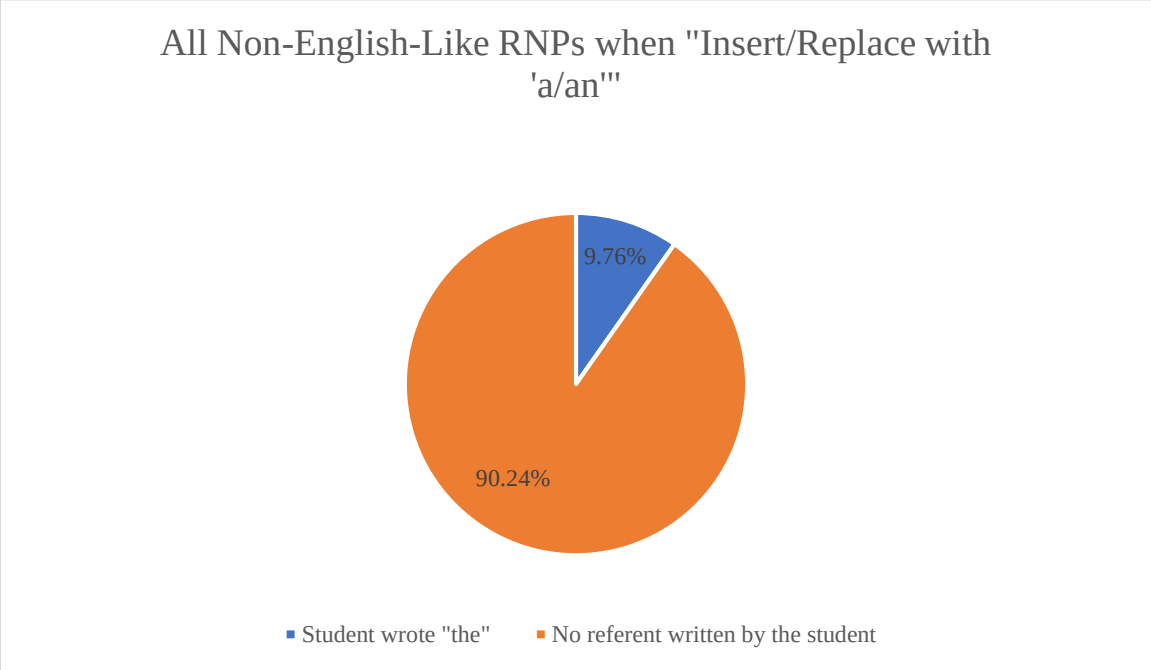

Figure 10: Non-English-like RNPs, for which “a/an” would be included in the English-like variation

For each instance when the evaluator deemed RNPs non-English-like, and she chose "Insert/Replace with 'the'" as the English-like variation, Figure 10 shows the proportion of each type of referent form written originally. In Essay 42 (below), RNP 46 and 47: “[46]culture and

[47]food"are examples of when "non-English-like" was chosen by the evaluator, and the writer had no article, referent form, or other word before each noun. As can be seen within the pie chart in Figure 4.7, of all the RNPs where the evaluator chose "non-English-like" and "Insert/Replace with 'the'", 87.5% of those RNPs had no article or other word given by the writer. For all the instances of RNPs where the English-like variation selected was "Insert/Replace with 'the'", 87.5% had no article or other word given before the noun. The deletion of "the" was a prominent non-English-like variation within the student essays which made up the data for this research.

Essay 42, sentence in second paragraph:

"for example, if [40]a new student decides to live in [41]university dormitory of [42]SCSU It will be easier for [43]him to know [44]different places within [45]the university, [46]culture and [47]food."

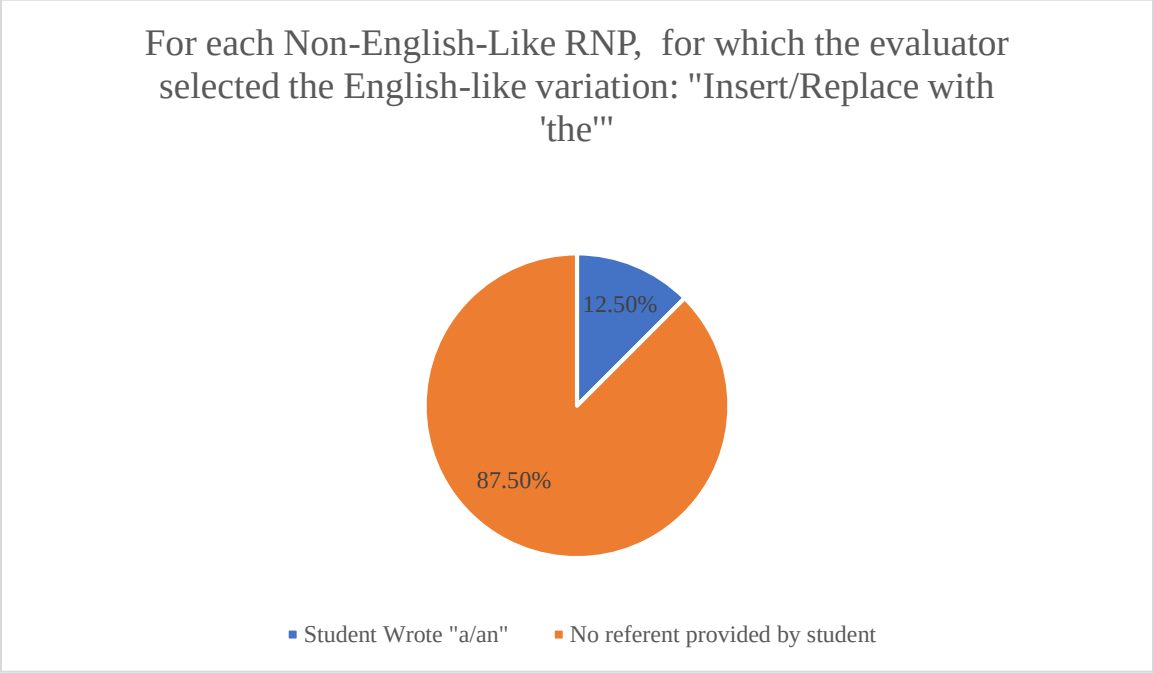

Figure 11: Non-English-like RNPs, for which "the" would be included in the English-like variation

In the corpus examined, the distribution of all of the RNPs across the Givenness Hierarchy, the English-likeness in each cognitive status category, and the frequent non-English-like variations of article use were researched. The distribution of dispersion of RNPs throughout the cognitive statuses recognized by the Givenness Hierarchy found the most-prevalent statuses to be those in the middle of the Hierarchy: Familiar and Uniquely Identifiable. The percent of English-like RNPs, compared with all the RNPs (English-like and non-English-like), resulted in high English-like precentages in all of the cognitive statuses, staying above 90%. The most frequemt non-English-like variations of article use were the deletion of “the” and “a”, and the oversuppliance of “the”. The distribution of the RNPs within the cognitive statuses of the Givenness Hierarchy answer the research question, and the other findings are secondarly results, demonstrating more detailed aspects of this group of students’ English article use.

Chapter V: Discussion

This research sought to understand how students who were from Nepal, and newly entering an American University, used English articles as classified by the Givenness Hierarchy. The data was made up of thirty student placement essays, which were written in response to a text-dependent prompt. Those essays were transcribed to a digital format, each referential noun phrase RNP was identified by the researcher. Each referential noun phrase (RNP) was evaluated in two ways: English-likeness and cognitive status. Firstly, an evaluator assessed the English-likeness of the article use within each RNP; for the non-English-like RNPs, she selected the English-like variation from the following choices: “Insert/Replace with 'the'", "Insert/Replace with ‘a/an’", and “Remove article provided by writer”. Secondly, the researcher recognized the cognitive status of each RNP with the categories, definitions, and guidelines provided above for the Givenness Hierarchy. The distribution of RNPs across the Givenness Hierarchy, the English-likeness, and the frequent non-English-like constructions were calculated.

The research question for this study is as follows: How do Nepali students, who are newly entering an American university use English articles, as classified by the Givenness Hierarchy? The distribution of RNPs across the Giveness Hierarchy was calculated for both English-like use and non-English-use. Additionally, the percent of English-like RNPs within each cognitive status category was calculated, demonstrating a high level of English-like article use. For each non-English-like RNP, the original article given by the student was compared with the English-like variation selection by the researcher, which demonstrated common non-English-like uses of articles by this group of students.

Figures 3 and 4 (see Chapter 4) show the proportion of each of the cognitive statuses, as recognized by the Givenness Hierarchy, as they appear in the data evaluated for this study.

Familiar is the most prevelent cognitive status, with 30.09% of all RNPs being Familiar. The least-represented cognitive status is Type Identifiable, with 8.64% of the total RNPs. Figure 7 a visual representation of the English-like percentage within each of the six cognitive statuses within the Givenness Hierarchy; the English-likeness percentage remains above 90% for each of the cognitive statuses. Figures 7 displays the prevalent oversuppliance of "the", when the evaluator's English-like variation had no article in that RNP. Figures 7 and 8 express the prevailing deletion of "a/an" and "the"; the student writers often did not use an article, when the English-like variation would have "a/an" for some RNPs, and "the" for others. The oversuppliance of "the" and the deletion of articles were frequent in the non-English-like RNPs.

One area in which the above research could potentially contribute to other fields is the area of Artificial Intelligence (AI), as AI involves essential work with cognitive statuses. According to Waldrop (1988), Allen Newell had a theory, claiming that "all cognition involves some form of problem-solving", and Newell's theory was used to create a program that would solve problems. Researchers developed a program called "Soar" (Waldrop, 1988), which sought to solve problems based on the input gained by experience with or exposure to numerous problems. As AI researchers create programs to solve various kinds of problems, from "medical diagnosis" (Waldrop, 1988) to "designing computer programs", their AI programs learn out of experience, to solve problems better in the future. Such a problem solving software could use the input of language, including the use of articles and other referents, to correctly use those articles and referents in the future. Such an AI program might use the cognitive statuses of referential phrases to better understand how to correctly use referential forms and phrases. A program such as Soar could solve the problem of which article to use, or the even more complex program of how a machine could use language the way humans use language.

Another field in which the above research on the use of cognitive statuses of referential noun phrases (RNPs) could be used is in the area of language learner assessment. When language learners enter a language learning program, their language needs to be assessed so that the student can be placed in a class level in which "the student will find material neither too easy nor too difficult but appropriately challenging" (Brown & Abeywickrama, 2010, p. 10). Creating such tests is difficult, and Carr (2011) recognized that no test will perfectly perform its task of appropriately placing students in language levels. Research of the cognitive statuses referenced by language learners may be beneficial for such testing.

As language learners use a language other than their first language, their language production is different from their L1, and different from their target language (TL). Selinker (1992) proposed the term "interlanguage" to describe the language that is "at least partially different from the native language and the target L2" (p. 395). He proposed that interlanguage was a language with "a linguistic system in its own right" (1992, p. 395), and, therefore, interlanguage ought to be studied as such (see the Analyzing Learner Language section of Chapter 2 for further explanation of Interlanguage). As language teachers evaluate learners' language, which is considered interlanguage, they may benefit from information about how the cognitive statuses are used by speakers to communicate.

Research of language learners' use of cognitive statuses as they communicate with referential forms (*a, the, this, that*) may greatly benefit language teachers in their evaluation of language learners. Research could be conducted, with participants from various language levels, and that data could be analyzed to see what differences may be evident between different language learning levels. It can be expected that language learners' interlanguage manifests different uses of cognitive statuses at different language proficiency levels; the differences

between levels could include accuracy of cognitive status use (as evidenced by referential form use) or a varying distribution of RNPs in each cognitive status recognized by the Givenness Hierarchy (Gundel et al., 1993). With such data available, language placement professionals could use the information about the use it to more accurately assess language learners' proficiency level.

The above research of the use of articles, as classified by the Givenness Hierarchy, may be utilized by AI research and language learner placement. As AI researchers create software to solve problems through ongoing exposure to problems being solved, this type of technology could seek to solve the problem of well-formed article use in machines that produce speech, or in language learners. Additionally, this research could contribute to the field of placement testing of language learners, if more research is done comparing the use of referential forms within varying cognitive statuses. However both of these potential contributions to other fields of study would also require more research on the use of referents on the Givenness Hierarchy Framework.

Chapter VI: Conclusion

Explanation of Findings

Through the evaluation of student written placement essays, the use of English articles by a specific group of students was studied. The data demonstrates a distribution across the cognitive statuses that are recognized by the Givenness Hierarchy; the most frequent cognitive statuses are in the center of the Giveness Hierarchy (Familiar and Uniquely Identifiable). The referents used by the students were 30.09% known to the reader either by some kind of shared knowledge, or mentioned at some point in the text; this qualified those RNPs as Familiar. There was a smaller proportion of cognitive statuses that were Type Identifiable (8.64%), which refer to a type of object, but not necessarily a specific object that is recognizable by the reader and the writer. What can be understood about the English-likeness of the students' article use is that they demonstrated a high proficiency of the use of English-like articles. Finally, an analysis of the non-English-like use of articles (or lack there of) revealed the common non-English-like constructions of article deletion and oversuppliance of "the". This research sought to understand the use of articles by this group of students, as defined by the Givenness Hierarchy.

Limitations of the Study

There were three limitations to this research: human error, difficulty in evaluation, and uncertainty of the student participants' first language (L1). Firstly, since human evaluators were used for the evaluation of English-likeness, the alternate English-like variation, identification of RNPs, and the identification of the cognitive statuses on the Givenness Hierarchy, human error was present. Secondly, the recognition of both RNPs and the cognitive status as recognized by the Givenness Hierarchy were both less-straightforward than expected. Thirdly, the data was collected for official University purposes before this research was conducted, therefore, the L1 of

the students is not known with certainty. Human error, difficulty in recognizing RNPs and cognitive status, and inability to know all the participants' L1 were the limitations of this research.

Since evaluations were completed using human evaluators, human error was a limitation of this study. As the grammaticality judgement evaluator was a human, it can be assumed that there was human error involved in the identification of the English-likeness of each RNP. As the researcher went through the grammaticality judgement evaluator's data, several errors were found. For example, the evaluator marked "remove article provided by writer", when there was no article provided by the writer. There were four of these instances, which were thrown out of the data. Additionally, as the researcher catalogued the decisions of the grammaticality judgement evaluator, she found instances of RNPs which were underlined, but not numbered in the data. Therefore, those RNPs were not included in the data, there were sixteen of these instances. Another area of human error found was in NPs that were identified as RNPs, but were later found to not be referential. Therefore, these phrases were thrown out of the data, thirty-three of these errors were found. Additionally, the researcher found other occurances of RNPs that ought to have been two RNPs, but were identified and numbered as one RNP. As these instances were not catalogued during the evaluation of the data, the researcher does not have accurate numbers on the number of these mis-identified RNPs. The researcher estimates that there was one mis-identified RNP per every second essay, totalling approximately fifteen of these RNPs. Since human evaluators were used for the assessment of this data, human error was a limitation.

Since the student essay data was collected by the university previous to this study (see the Materials section of Chapter 3), and the students were not contacted for the purposes of this research, some specific information will be unattainable. The students' L1 could not be known

with certainty. However, according to the Central Intelligence Agency (2018), nearly half of Nepali people consider Nepali their L1, and nearly two-thirds of Nepali people speak a language of Indo-Aryan descent (The Editors of Encyclopaedia Britannica, 2018). The lack of ability to know some specific information from the participants is a known limitation of this study.

As is explained above in the Analysis section of Chapter 3, both identifying RNPs and identifying the cognitive status of each RNP on the Givenness Hierarchy were, at times, complex decisions. The difficulty in identifying which NPs were Referential, and which were not was a limitation. Additionally, the close connection between some of the cognitive statuses within the Givenness Hierarchy, and difficulty in recognizing the correct cognitive status with certainty is another limitation. The challenge in determining the referential nature and the cognitive status of NPs are limitations of this research.

Three limitations to this research have been recognized: human error, limited knowledge of student participants, and difficulties in identifying RNPs and each cognitive status. Human error is unavoidable in this type of study; since human evaluators were used, human errror is assumed. Since the data already existed before this research was conducted, limited information was available about each student. Finally, recognizing RNPs from non-referential NPs proved to be difficult and not straighforward, as did the identification of the RNP of each cognitive status. These are known limitations of this research.

Suggestions for future research

Subsequent research could be completed with a different group of students. Written essays of students from Nepal were utilized for this research; future study could analyize a group of students from another country or language. This group could potentially be more linguistically homogenous. Also, future study could include students with a greater difficulty with English

articles. Research of an alternate group of students could be compared with the data from these students from Nepal; also, participants could have a higher percentage of non-English-like articles.

This research was completed using data from a group of students who are from Nepal, but future research could performed using students from another country or language. Another similar study could evaluate the English article use of students from another country. Since Nepal is linguistically diverse (see Chapter 1), a more linguistically homogenous group of students could be used as participants. Additionally, a group of students from another country could be compared to these Nepali students' English article use. Further research could be done by utilizing an alternate group of students.

Though the use of English articles is widely known to be a notable difficulty for most English learners (see the Classification Systems for English Articles section of Chapter 2), the student writing evaluated for this study do not appear to have a significant difficulty with English article use (see Figure 7). Future research could be done with a group of students either with a more significant difficult with English articles, or with a group of students with a wide range of difficulty with English article use. Further investigation could be done using participants with a significant level of difficulty with the use of English articles.

Further investigation can be completed by utilizing alternate groups of students. This research was conducted using students who were newly entering an American University, who had completed the University's proficiency requirements, and who were from Nepal. A similar group of participants could be researched, but who were from a different country, or specific language or language group could be studied. Alternatively, since this group of students had a high level of English-like use of English articles, participants with a lower level of proficiency,

including a lower proficiency with English article use, could be researched. By drawing participants from a disparate group of students could be helpful for further understanding in this area of study.

References

Al-Qadi, M. M. (2017). English Article System Errors among Saudi Arab EFL Learners: A Case of the Preparatory Year English Program Learners of King Saud University. *International Journal of English Language and Translation Studies*, *5*(01), 69-77.

Ariel, M. (1991). The Function of Accessibility in a Theory of Grammar. *Journal of Pragmatics*, *16*(5), 443–463.

Brown, H. D. & Abeywickrama, P. (2010). *Language Assessment Principles and Classroom Practices* (2nd Ed.). White Plains, NY: Pearson Education.

Carr, N. T. (2011). *Designing and Analyzing Language Tests*. Oxford, UK: Oxford University Press.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). The Grammar Book: An ESL/EFL Teacher's Course (2nd Ed.). Heinle & Heinle Publishers.

Central Intelligence Agency. (2018). CIA. Retrieved October 28, 2018, from: https://www.cia.gov

Corder, S. P. (1967). Significance of Learner's Errors. *IRAL - International Review of Applied Linguistics in Language Teaching*, *5*(1-4), 161-170.

Dean, A. C. (2017). Complex Dynamic Systems and Interlanguage Variability: Investigating Topic, Syntactic Complexity, and Accuracy in NS-NNS Written Interaction. *Working Papers in TESOL & Applied Linguistics*, *17*(1), 56–97.

The Editors of Encyclopaedia Britannica. (2018). Encyclopaedia Britannica, Inc. Retrieved October 28, 2018, from: https://www.britannica.com/

Ekiert, M. (2016). *Article omission: Toward establishing how referents are tracked in L2 English*. Georgetown University Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language: Journal Of The Linguistic Society Of America, 2*(2), 274-307.

Ishaq, Q., & Bukhari, N. H. (2016). Persistent Errors in English Writing and speech Made by Urdu Undergraduates at UAJK: A Case of Mother Tongue Interference. *Kashmir Journal of Language Research*, *19*(1), 249–264.

Lakshmanan, U., & Selinker, L. (2001). Analysing interlanguage: how do we know what learners know? *Second Language Research*, *17*(4), 393–420.

Leonhardt, D. (2018). A College Experiment That Really Worked. *The New York Times.* Retrieved from: https://www.nytimes.com/2018/12/11/opinion/college-gap-michigan-university.html?rref=collection%2Fsectioncollection%2Fopinion-columnists

Lightbown, P.M. & Spada, N. (2013). How Languages are Learned. Oxford, UK: Oxford University Press.

Master, P. (1990). Teaching the English Articles as a Binary System. *TESOL Quarterly*, (3), 461.

Richards, J. C. (1970). A Non-Contrastive Approach to Error Analysis.

Sadrai, E. (2016). Referents of *ra*-Marked Nominal Expressions in Persian Discourse. Unpublished

Selinker, L. (1969). Language Transfer. *General Linguistics*, 67–92.

Selinker, L. Rediscovering Interlanguage. (1992). New York, NY: Longman Group UK Limited 1992.

Simons, G. F. & Fennig, C. D. (eds.). (2018). *Ethnologue: Languages of the World, Twenty-first edition.* Dallas, Texas: SIL International. Retrieved November 5, 2018 from: http://www.ethnologue.com.

Summer Institute of Linguistics., Gordon, R. G., & Grimes, B. F. (2005). *Ethnologue: Languages of the world* (15th ed.). Dallas, Tex.: SIL International.

Swierzbin, B. J. (2004). The Role of Cognitive Status in Second Language Acquisition of English Noun Phrase Referring Expressions. *Dissertation Abstracts International, Section A: The Humanities and Social Sciences*, *6*(6).

Tarone, E., Parrish, B. (1988). Task-Related Variation in Interlanguage: The Case of Articles. In Guiora, A. Z. (1988). *Language Learning. 38* (1). (21-44).

Waldrop, M. (1988). Toward a unified theory of cognition: Soar is a computer program that can learn from experience - and that may also explain the basic mechanisms of thought. (cognitive science and artificial intelligence, part 1). *Science*, *241*(4861), 27–29.

Zhao, H., & Macwhinney, B. (2018). The Instructed Learning of Form-Function Mappings in the English Article System. *Modern Language Journal*, (1), 99.

Appendix A: IRB Approval Form

ST. CLOUD STATE UNIVERSITY
EDUCATION FOR LIFE.

Institutional Review Board (IRB)
720 4th Avenue South AS 210, St. Cloud, MN 56301-4498

Name: Elizabeth Claessens

Email: eclaessense@stcloudstate.edu

IRB PROTOCOL DETERMINATION: Exempt Review

Project Title: Article Use of English Learners Classified by the Giveness Hierarychy

Advisor Edward Sadrai

The Institutional Review Board has reviewed your protocol to conduct research involving human subjects. Your project has been: **APPROVED**

Please note the following important information concerning IRB projects:

- The principal investigator assumes the responsibilities for the protection of participants in this project. Any adverse events must be reported to the IRB as soon as possible (ex. research related injuries, harmful outcomes, significant withdrawal of subject population, etc.).

- For expedited or full board review, the principal investigator must submit a Continuing Review/Final Report form in advance of the expiration date indicated on this letter to report conclusion of the research or request an extension.

-Exempt review only requires the submission of a Continuing Review/Final Report form in advance of the expiration date indicated in this letter if an extension of time is needed.

- Approved consent forms display the official IRB stamp which documents approval and expiration dates. If a renewal is requested and approved, new consent forms will be officially stamped and reflect the new approval and expiration dates.

- The principal investigator must seek approval for any changes to the study (ex. research design, consent process, survey/interview instruments, funding source, etc.). The IRB reserves the right to review the research at any time.

If we can be of further assistance, feel free to contact the IRB at 320-308-4932 or email ResearchNow@stcloudstate.edu and please reference the SCSU IRB number when corresponding.

IRB Chair:

Dr. Benjamin Witts
Associate Professor- Applied Behavior Analysis
Department of Community Psychology,Counseling, and Family Therapy

IRB Institutional Official:

Dr. Latha Ramakrishnan
Interim Associate Provost for Research
Dean of Graduate Studies

OFFICE USE ONLY

SCSU IRB# 1876 - 2408	**Type:** Exempt Review	**Today's Date:** 2/6/2019
1st Year Approval Date: 2/5/2019	**2nd Year Approval Date:**	**3rd Year Approval Date:**
1st Year Expiration Date:	**2nd Year Expiration Date:**	**3rd Year Expiration Date:**

Appendix B: Directions for Grammaticality Judgement Evaluator

Thank you for spending your time to help me do some research for my thesis.
I am looking for an English-speaker's (specifically, native English monolingual (never functioned in another language) and non-linguistic-expert) view of article use. Below is an example passage. Specific words and prhases have been highlighted and numbered.

Directions: Based on your first instinct, decide whether the article use (only) in each word/phrase is English-like or non-English-like. Think: "Would a native speaker of English (somewhere) use the article or lack of article in this way?" If the answer is yes, choose the "English like article" box. If the answer is no, choose what you think the correct article use should be: "insert/replace with 'the'", "insert/replace with 'a/an'", or "remove article provided by writer".

"It's fashionable to question [1]the value of [2]a four-year college degree. But [3]three basic facts continue to make clear how valuable such [4]the degree is: One, [5]college graduates fare better by [6]virtually every available metric — [7]income, [8]wealth, [9]health, [10]life satisfaction and more. Two, [11]a careful studies suggest that [12]the college plays [13]a causal role in improving [14]people's lives. And, three, [15]virtually everyone with [16]the ability to send [17]their own children to [18]college — including [19]people who are publicly skeptical of [20]education — does so. Unfortunately, [21]most working-class and poor teenagers, including [22]many who excel in [23]high school, still don't graduate from [24]college. [25]They often enroll in [26]colleges that have [27]a high dropout rate and never finish." (Leonhardt, 2018)

1	English-like article	Non-English article		
		Insert/replace with "the"	Insert/replace with "a/an"	Remove article provided by writer
2	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
3	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
4	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
5	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
6	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
7	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
8	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
9	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer
10	English-like article	Non-English article		
		Insert/replace with "the"	Replace with "a/an"	Remove article provided by writer

11	English-like article	Non-English article		
		Insert/replace with “the”	Insert/replace with “a/an”	Remove article provided by writer
12	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
13	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
14	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
15	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
16	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
17	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
18	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
19	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
20	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer

21	English-like article	Non-English article		
		Insert/replace with “the”	Insert/replace with “a/an”	Remove article provided by writer
22	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
23	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
24	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
25	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
26	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
27	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer

Appendix C: Evaluative Form

1	English-like article	Non-English article		
		Insert/replace with “the”	Insert/replace with “a/an”	Remove article provided by writer
2	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
3	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
4	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
5	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
6	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
7	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
8	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
9	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer
10	English-like article	Non-English article		
		Insert/replace with “the”	Replace with “a/an”	Remove article provided by writer

Appendix D: Data Table

Number	Code	Word Count	Total RNPs	Cognitive Status	Total Number (Percent of Whole)	Number English-like (Percent English-like)	Number Non-English-like
1	3	224	57	In Focus	8 (14%)	8 (100%)	0
				Activated	7 (12%)	7 (100%)	0
				Familiar	16 (28%)	16 (100%)	0
				Uniquely Identifiable	11 (19%)	10 (91%)	1 (9%)
				Referential	7 (12%)	5 (71%)	2 (29%)
				Type Identifiable	8 (14%)	7 (88%)	1 (13%)
2	5	304	85	In Focus	10 (12%)	10 (100%)	0
				Activated	19 (22%)	19 (100%)	0
				Familiar	23 (27%)	22 (96%)	1 (4%)
				Uniquely Identifiable	13 (15%)	13 (100%)	0
				Referential	12 (14%)	10 (83%)	2 (17%)
				Type Identifiable	8 (9%)	7 (88%)	1 (13%)
3	7	271	75	In Focus	18 (24%)	18 (100%)	0
				Activated	8 (11%)	7 (88%)	1 (13%)
				Familiar	10 (13%)	9 (90%)	1 (10%)
				Uniquely Identifiable	10 (13%)	8 (80%)	2 (20%)
				Referential	19 (25%)	18 (95%)	1 (5%)
				Type Identifiable	10 (13%)	10 (100%)	0
4	8	419	117	In Focus	12 (10%)	12 (100%)	0
				Activated	20 (17%)	20 (100%)	0
				Familiar	31 (26%)	31 (100%)	0
				Uniquely Identifiable	18 (15%)	18 (100%)	0
				Referential	25 (21%)	24 (96%)	1 (4%)
				Type Identifiable	11 (9%)	10 (91%)	1 (9%)
5	11	432	100	In Focus	10 (10%)	10 (100%)	0
				Activated	6 (6%)	6 (100%)	0
				Familiar	30 (30%)	25 (83%)	5 (17%)
				Uniquely Identifiable	24 (24%)	21 (88%)	3 (13%)
				Referential	18 (18%)	15 (83%)	3 (17%)
				Type Identifiable	12 (12%)	11 (92%)	1 (8%)
6	12	311	78	In Focus	4 (5%)	4 (100%)	0
				Activated	6 (8%)	6 (100%)	0
				Familiar	25 (32%)	25 (100%)	0
				Uniquely Identifiable	26 (33%)	22 (85%)	4 (15%)
				Referential	13 (17%)	13 (100%)	0
				Type Identifiable	4 (5%)	3 (75%)	1 (25%)
7	13	424	111	In Focus	21 (19%)	21 (100%)	0
				Activated	16 (14%)	16 (100%)	0
				Familiar	21 (19%)	21 (100%)	0
				Uniquely Identifiable	30 (27%)	30 (100%)	0
				Referential	17 (15%)	15 (88%)	2 (12%)
				Type Identifiable	6 (5%)	6 (100%)	0
8	14	302		In Focus	3 (4%)	3 (100%)	0

			79	Activated	18 (22%)	18 (100%)	0
				Familiar	27 (33%)	26 (96%)	1 (4%)
				Uniquely Identifiable	15 (19%)	14 (93%)	1 (7%)
				Referential	4 (5%)	4 (100%)	0
				Type Identifiable	14 (17%)	14 (100%)	0
9	15	273	87	In Focus	16 (18%)	16 (100%)	0
				Activated	12 (14%)	11 (92%)	1 (8%)
				Familiar	24 (28%)	11 (92%)	2 (8%)
				Uniquely Identifiable	20 (23%)	18 (2%)	2 (10%)
				Referential	8 (9%)	8 (100%)	0
				Type Identifiable	7 (8%)	7 (100%)	0
10	16	296	89	In Focus	18 (20%)	16 (89%)	2 (11%)
				Activated	11 (12%)	8 (73%)	3 (27%)
				Familiar	29 (33%)	27 (93%)	2 (7%)
				Uniquely Identifiable	15 (17%)	15 (100%)	0
				Referential	6 (7%)	6 (100%)	0
				Type Identifiable	10 (11%)	9 (%)	1 (10%)
11	17	313	84	In Focus	11 (13%)	10 (91%)	1 (92%)
				Activated	13 (15%)	12 (92%)	1 (8%)
				Familiar	17 (20%)	17 (100%)	0
				Uniquely Identifiable	29 (35%)	28 (97%)	1 (3%)
				Referential	7 (8%)	7 (100%)	0
				Type Identifiable	7 (8%)	7 (100%)	0
12	18	327	82	In Focus	3 (4%)	3 (100%)	0
				Activated	16 (20%)	16 (100%)	0
				Familiar	21 (26%)	19 (90%)	2 (10%)
				Uniquely Identifiable	19 (23%)	17 (89%)	2 (11%)
				Referential	16 (20%)	13 (81%)	3 (19%)
				Type Identifiable	7 (9%)	7 (100%)	0
13	21	291	72	In Focus	13 (18%)	13 (100%)	0
				Activated	8 (11%)	8 (100%)	0
				Familiar	23 (32%)	22 (96%)	1
				Uniquely Identifiable	18 (25%)	18 (100%)	0
				Referential	8 (11%)	8 (100%)	0
				Type Identifiable	2 (3%)	2 (100%)	0
14	23	345	99	In Focus	7 (7%)	7 (100%)	0
				Activated	19 (19%)	19 (100%)	0
				Familiar	35 (35%)	35 (100%)	0
				Uniquely Identifiable	21 (21%)	20 (95%)	1 (5%)
				Referential	6 (6%)	6 (100%)	0
				Type Identifiable	6 (6%)	6 (100%)	0
15	24	506	134	In Focus	19 (14%)	18 (95%)	1 (5%)
				Activated	23 (17%)	23 (100%)	0
				Familiar	49 (37%)	47 (96%)	2 (%)
				Uniquely Identifiable	35 (26%)	34 (97%)	1 (3%)
				Referential	7 (5%)	6 (86%)	1 (14%)
				Type Identifiable	1 (1%)	1 (100%)	0
16	25	566		In Focus	21 (14%)	20 (95%)	1 (5%)

			146	Activated	21 (14%)	21 (100%)	0
				Familiar	46 (32%)	43 (93%)	3 (7%)
				Uniquely Identifiable	40 (34%)	34 (85%)	6 (15%)
				Referential	11 (8%)	8 (73%)	3 (27%)
				Type Identifiable	7 (5%)	7 (100%)	0
17	26	353	101	In Focus	9 (9%)	8 (89%)	1 (11%)
				Activated	14 (14%)	14 (100%)	0
				Familiar	33 (33%)	32 (97%)	1 (3%)
				Uniquely Identifiable	27 (27%)	26 (96%)	1 (4%)
				Referential	8 (8%)	8 (100%)	0
				Type Identifiable	10 (10%)	10 (100%)	0
18	27	382	104	In Focus	17 (16%)	17 (100%)	0
				Activated	25 (24%)	24 (96%)	1 (4%)
				Familiar	15 (14%)	13 (87%)	2 (13%)
				Uniquely Identifiable	20 (19%)	16 (80%)	4 (20%)
				Referential	13 (13%)	12 (92%)	1 (8%)
				Type Identifiable	14 (13%)	14 (100%)	0
19	28	477	139	In Focus	21 (15%)	21 (100%)	0
				Activated	25 (18%)	23 (92%)	2 (8%)
				Familiar	39 (28%)	36 (92%)	3 (8%)
				Uniquely Identifiable	14 (10%)	14 (100%)	0
				Referential	15 (11%)	12 (80%)	3 (20%)
				Type Identifiable	25 (18%)	23 (92%)	2 (8%)
20	29	370	103	In Focus	10 (10%)	10 (100%)	0
				Activated	14 (14%)	12 (86%)	2 (14%)
				Familiar	32 (31%)	30 (94%)	2 (6%)
				Uniquely Identifiable	30 (29%)	29 (100%)	0
				Referential	10 (10%)	10 (100%)	0
				Type Identifiable	7 (7%)	5 (71%)	2 (29%)
21	30	345	96	In Focus	11 (11%)	11 (100%)	0
				Activated	17 (18%)	16 (94%)	1 (6%)
				Familiar	29 (30%)	23 (79%)	6 (21%)
				Uniquely Identifiable	28 (29%)	23 (82%)	5 (18%)
				Referential	7 (7%)	6 (86%)	1 (14%)
				Type Identifiable	4 (4%)	4 (100%)	0
22	31	405	111	In Focus	13 (12%)	13 (100%)	0
				Activated	27 (24%)	25 (93%)	2 (7%)
				Familiar	46 (41%)	44 (96%)	2 (4%)
				Uniquely Identifiable	20 (18%)	20 (100%)	0
				Referential	4 (4%)	4 (100%)	0
				Type Identifiable	1 (1%)	1 (100%)	0
23	33	399	85	In Focus	4 (5%)	4 (100%)	0
				Activated	13 (15%)	12 (92%)	1 (8%)
				Familiar	31 (36%)	31 (100%)	0
				Uniquely Identifiable	15 (18%)	14 (93%)	1 (7%)
				Referential	12 (14%)	12 (100%)	0
				Type Identifiable	10 (12%)	10 (100%)	0
24	34	278		In Focus	13 (15%)	13 (100%)	0

			85	Activated	17 (20%)	17 (100%)	0
				Familiar	22 (26%)	21 (95%)	1 (5%)
				Uniquely Identifiable	24 (28%)	23 (96%)	1 (4%)
				Referential	1 (1%)	1 (100%)	0
				Type Identifiable	8 (9%)	8 (100%)	0
25	40	297	65	In Focus	6 (9%)	6 (100%)	0
				Activated	11 (17%)	9 (82%)	2 (18%)
				Familiar	30 (46%)	29 (97%)	1 (3%)
				Uniquely Identifiable	12 (18%)	12 (100%)	0
				Referential	4 (6%)	4 (100%)	0
				Type Identifiable	2 (3%)	2 (100%)	0
26	42	361	107	In Focus	10 (9%)	10 (100%)	0
				Activated	22 (21%)	21 (95%)	1 (5%)
				Familiar	31 (29%)	27 (87%)	4 (13%)
				Uniquely Identifiable	19 (18%)	16 (84%)	3 (16%)
				Referential	11 (10%)	10 (91%)	1 (9%)
				Type Identifiable	14 (13%)	12 (86%)	2 (14%)
27	46	430	79	In Focus	3 (4%)	3 (100%)	0
				Activated	6 (8%)	4 (67%)	1 (33%)
				Familiar	28 (35%)	27 (96%)	1 (4%)
				Uniquely Identifiable	24 (30%)	24 (100%)	0
				Referential	6 (8%)	6 (100%)	0
				Type Identifiable	12 (15%)	12 (100%)	0
28	47	600	166	In Focus	20 (12%)	20 (100%)	0
				Activated	30 (18%)	29 (97%)	1 (3%)
				Familiar	63 (38%)	56 (89%)	7 (11%)
				Uniquely Identifiable	38 (23%)	37 (97%)	1 (3%)
				Referential	11 (7%)	10 (91%)	1 (9%)
				Type Identifiable	4 (2%)	4 (100%)	0
29	48	567	162	In Focus	33 (20%)	33 (100%)	0
				Activated	28 (17%)	28 (100%)	0
				Familiar	42 (26%)	41 (98%)	1 (2%)
				Uniquely Identifiable	30 (19%)	28 (93%)	2 (7%)
				Referential	12 (7%)	11 (92%)	1 (8%)
				Type Identifiable	17 (10%)	17 (100%)	0
30	49	373	93	In Focus	12 (13%)	12 (100%)	0
				Activated	10 (11%)	10 (100%)	0
				Familiar	32 (34%)	30 (94%)	2 (6%)
				Uniquely Identifiable	21 (23%)	21 (100%)	0
				Referential	11 (12%)	11 (100%)	0
				Type Identifiable	7 (8%)	7 (100%)	0