

# **Tagalog-English code-switching: issues in the nominal domain**

Ramon Lorenzo D. Labitigan

Senior Thesis Advisor: Raffaella Zanuttini

Submitted to the faculty of the Department of Linguistics  
in partial fulfillment of the requirements for the degree of

Bachelor of Arts

Yale University

May 1, 2013

# Table of Contents

<b>Acknowledgments.....</b>	<b>4</b>
<b>0 Abstract.....</b>	<b>5</b>
<b>1 Introduction.....</b>	<b>6</b>
1.1 Code-switching as a research topic	6
1.2 Tagalog-English bilingualism	6
1.2.1 Tagalog	6
1.2.2 English in Tagalog-speaking populations	7
1.2.3 Tagalog-English code-switching (TECS)	7
1.3 Opening remarks	7
<b>2 Theoretical framework.....</b>	<b>8</b>
2.1 The Matrix Language Frame (MLF) Model	8
2.1.1 The ML:EL opposition	9
2.1.2. The content-system morpheme opposition	11
2.2 The Abstract Level Model	12
2.3 Congruence	13
2.4 The 4-Morpheme (4-M) Model	14
2.5 Summary	15
<b>3 Plural markers in Tagalog, English, and TECS.....</b>	<b>19</b>
3.1 Plural markers in Tagalog and English	19
3.2 Plural markers in TECS	21
3.2.1 Observations	21
3.2.1.1 English as the ML	21
3.2.1.2 Tagalog as the ML	21

3.2.2 Analysis	22
3.2.2.1 English as the ML	22
3.2.2.2 Tagalog as the ML	23
3.2.3 Discussion	27
3.2.3.1 Sufficient congruence is required for acceptable code-switches	27
3.2.3.2 Lemmas are language-specific	29
<b>4 Subjecthood and case in Tagalog, English, and TECS.....</b>	<b>29</b>
4.1 Subjecthood and case in Tagalog and English	30
4.1.1 The Tagalog topic system	30
4.1.2 The puzzle of subjecthood in Tagalog	31
4.2 Subjecthood and case in TECS	33
4.2.1 Observations	33
4.2.1.1 Tagalog as the ML	33
4.2.1.2 English as the ML	34
4.2.2 Analysis	35
4.2.2.1 Tagalog as the ML	35
4.2.2.2 English as the ML	35
4.2.3 Discussion	38
4.2.3.1 Code-switching can inform monolingual analyses	38
4.2.3.2 A puzzling exception to the “primacy of the <i>ang</i> -marked form”	38
<b>5 Conclusion.....</b>	<b>40</b>
5.1 Summary	40
5.2 Future directions	41
<b>References.....</b>	<b>42</b>

## Acknowledgments

This senior thesis has been quite a remarkable journey. It began with an interest in a topic that I had never studied before, but had been exposed to all my life: Tagalog-English code-switching (TECS). Wading through literature on your own is a vastly intimidating task, but the experience has been unbelievably rewarding. The learning curve was steep, the nights long, and the frustration abundant, but thanks to the unfailingly supportive community here at the Yale Linguistics Department, I was able to stick with the topic and slowly develop it into a final product

In particular, I would like to thank Professor Raffaella Zanuttini for always being calm, patient, and helpful and for always caring about the work and growth of the undergraduates. She was my first contact in the Linguistics Department and has remained one of my most supportive mentors ever since.

And of course, I would like to thank the record-breaking linguistics majors of the Yale Class of 2013. One of the most fun and weird communities on campus, you have been one of the staples of my Yale career. Thanks for all the support and the endless, endless wordplay. And remember, when in doubt, just do linguistics on it.

## **0 Abstract**

Research on code-switching, the use of two or more languages in a single utterance, tends to focus on either the structural factors or the social factors that govern what code-switches are acceptable and unacceptable. The Matrix Language Frame (MLF) Model (Myers-Scotton, 2002) explains the structural aspects of code-switching by acknowledging an asymmetry between participating languages. The model designates a matrix language (ML) and an embedded language (EL) for a given CP.

This paper investigates two structural aspects of code-switched nominal phrases in bilingual Tagalog-English speech, namely nominal pluralization and case/subjecthood, using a theoretical framework consisting of the MLF Model and other supporting models. Tagalog-English code-switching (TECS) is widely accepted among bilingual speakers in the Philippines and communities around the world. Despite the prevalence of this phenomenon, there has been limited work focusing on this language pair in the code-switching literature. The code-switching literature has been dominated by language pairs that are typologically similar. Further study on typologically dissimilar pairs such as Tagalog and English will be extremely valuable in understanding the mechanisms underlying code-switching.

# **1 Introduction**

## **1.1 Code-switching as a research topic**

Bilingualism, as an area of study, has sparked great interest across many fields, ranging from linguistics and cognitive science to sociology and politics. The study of bilingualism has many implications for linguistics, as insights gained may support current models of linguistic theory, refute them, or provide novel insights that cannot be verified from studying monolingual speakers alone.

Code-switching, one particular phenomenon of bilingual speech, refers to instances of alternating between two languages or varieties of the same language in the same conversation (Myers Scotton, 1983). The linguistic research on code-switching can be generally grouped according to two approaches: structural and sociolinguistic. The structural approach seeks to characterize how code-switching can be represented in the mind. The sociolinguistic approach views code-switching as a sociopragmatic phenomenon, focusing on the social motivations and functions of code-switching (Amuda, 1994). Although both of these main perspectives are invaluable in order to fully understand code-switching, this paper focuses on the grammatical structure of code-switching.

## **1.2. Tagalog-English bilingualism**

### **1.2.1. Tagalog**

Tagalog is the Philippines' most widely spoken language. Filipino, the officially recognized national language of the Philippines, is based chiefly on Tagalog and can be considered a variety of Tagalog. Tagalog serves as the lingua franca throughout the country and throughout the world in communities of the Philippine diaspora. Tagalog has 28 million native speakers in the Philippines, but almost the entire country's population of 95 million can speak it as a first, second, or third language (Lewis et al., 2013). The dominance of Tagalog was achieved largely through the educational system

and mass media. Roughly 10 million Tagalog speakers are part of the global Philippine diaspora, 1.5 million of which live in the United States. Tagalog is the fourth most widely spoken language in the United States, after English, Spanish, and Chinese languages (mostly Cantonese) (Shin et al., 2010).

### **1.2.2. English in Tagalog-speaking populations**

English, the second national language of the Philippines, took root during a period of colonization by the United States starting in 1898 at the end of the Spanish-American War. In the Philippines, English is now widely spoken, particularly by higher socioeconomic groups, and maintains a central role in government, business, technology, and mass media. English is also used widespread as a lingua franca among the international Filipino community.

### **1.2.3. Tagalog-English code-switching (TECS)**

Tagalog-English code-switching (TECS), or Taglish, is a variety of bilingual speech. Although it can be considered a prestige language variety, TECS has an increasingly substantial presence in all socioeconomic classes in the Philippines, particularly in urban centers. TECS is also widely spoken by bilingual communities throughout the world.

## **1.3 Opening remarks**

Despite its widespread use and the positive attitudes toward this style of speech, TECS is underrepresented in the code-switching literature. This paper takes one small step at filling this gap.

It is well known among linguists that nouns are the most code-switched or borrowed forms. In the case of TECS, the nominal domain, which is the subject of this paper, provides many interesting phenomena that require explanation. For the remainder of the paper, I will often be using the term *nominal phrase* when referring to constituents in the nominal domain, that is, constituents headed by an N. *Nominal phrase* is a non-committal term that helps our analysis in two ways. First, *nominal phrases* of English and of Tagalog seem to behave very differently. Nominal phrases in Tagalog remain a

challenging topic of study, and there still remain many competing ideas about their structure and properties. Thus, a general term such as *nominal phrase* when referring to Tagalog relieves our analysis of some unnecessary complexity. Second, different types of English nominal phrases (i.e. Ns, modified NPs, conjoined NPs, DPs) seem to sometimes pattern together in TECS (for example, see section 3.2.3.1). Thus, an umbrella term helps to capture certain patterns in the data.

This paper examines two areas of morphosyntax within the nominal domain. The first, involving nominal plural markers, can be considered an aspect of the internal syntax of nominal phrases. The second, involving issues such as subjecthood and case, can be considered a fragment of the external syntax of nominal phrases, which has direct relevance for the greater structure of the CP.

## **2 Theoretical framework**

Many of the early attempts (from the 1970s into the 1990s) to explain the grammatical structure of code-switching were constraint-based descriptive approaches. Constraints were empirically motivated, but not placed into the context of any specific theoretical model or approach.

### **2.1 The Matrix Language Frame (MLF) Model**

Unlike many other approaches to code-switching, the Matrix Language Frame (MLF) Model (Myers-Scotton, 1993 [1997]) is not simply a collection of descriptive constraints. Rather, it is a multi-layer model with interconnected parts that not only describes linguistic phenomena, but also provides an explanation for why these phenomena occur.

At the core of the MLF Model are two key oppositions based on asymmetries in code-switching structures: the Matrix Language (ML) – Embedded Language (EL) opposition and the content-system morpheme opposition.



### 2.1.1 The ML:EL opposition

The first opposition stems from the fact that the languages involved in code-switching do not participate equally. A higher level of participation does not refer to a greater number of morphemes or even the presence of certain morphemes, but rather the contribution of more *abstract structure*. The language that contributes more abstract structure can be referred to as the Matrix Language (ML), while the other language can be referred to as the Embedded Language (EL).

The unit of analysis of the MLF Model is the CP (projection of complementizer). A CP is the highest projection of the clause. This unit of analysis for the MLF Model does not only account for the data regarding distributions of the two participating languages, but it also provides an easily identifiable and consistent unit for comparisons across examples and languages. Referring to the CP also allows us to avoid the technical difficulty in defining and distinguishing among other terms such as sentence, clause, and utterance in our analysis.

For each CP, there is a grammatical frame specified. This frame, called the ML, is abstract in nature; it does not itself include any actual morphemes, but rather, “it includes specifications about slots and how they are to be filled, based on directions from lemmas in the mental lexicon” (Myers-Scotton, 2002, p. 67). There is quantitative evidence suggesting that the ML cannot switch within a CP (Finlayson et al., 1998), making the CP the appropriate atomic structure for the study of code-switching.

In monolingual speech, the ML frame of each CP is “vacuously transparent” (Jake et al., 2002, p. 72) since the frame is provided by the speaker’s only language. In bilingual speech, this frame may be provided by either one of the two participating languages or, in certain types of contact phenomena, by a combination of the two. For the purposes of this paper, we will only consider the case when only one language provides the grammatical frame for a CP. It is important to note that although we will be using “ML” as a label for the language from which the grammatical frame is abstracted, the ML itself, i.e. the ML frame, is not actually synonymous with either of the participating languages. That is, the frame is “not its source language, but rather represents an abstraction from it” (Myers-Scotton, 2002, p. 67).

The ML:EL distinction thus allows for three types of constituents in code-switching speech. The first two are ML islands and EL islands, which are constituents made up completely of one language and are well-formed in that language. We can also define islands as referring to constituents that have dependency relations, and thus must be comprised of two or more morphemes. ML islands do not pose any issues, as they behave as they would in monolingual speech. On the other hand, “EL islands represent a break in the ML frame” (Myers-Scotton, 2002, p. 139). Myers-Scotton (2002) argues that this does not necessarily pose a problem for the MLF Model, but rather calls to attention the complexity behind the relative levels of activation of participating languages. The details of this justification will not be explored here.

The third type of constituent is a mixed ML-EL constituent, which may contain iterations of islands and other mixed constituents. This means that we can consider the bilingual CP with at least one code-switch to be a large mixed constituent. A mixed constituent, aside from containing an EL island, can also contain a singly-occurring EL form. Whereas an EL island is a constituent made from multiple morphemes, a singly-occurring EL form contains one content morpheme (to be defined in section 2.1.2). There continues to be debate about whether singly-occurring forms are best analyzed as single-word code-switches or borrowings. Borrowings are often classified as either established borrowings or unestablished “nonce” borrowings. Myers-Scotton (2002) argues that, with the MLF Model, there is no need to make a distinction between code-switches and borrowings. The MLF Model accounts for all singly occurring EL forms; that is, borrowings and singly-occurring code-switched forms (and perhaps even intermediate stages, if they were to exist) “largely are integrated into the morphosyntactic frame of the recipient or [ML]” (Myers-Scotton, 2002, p. 153). In fact, this “integration” of forms into the ML frame even occurs in monolingual speech, as mentioned earlier. In terms of borrowings and singly-occurring code-switch, the same principles regarding a requirement for “sufficient congruence” apply (to be to be discussed in section 2.3). The only distinction between a singly-occurring code-switch and a borrowing that is relevant from the MLF perspective may be their status in the mental lexicon. Thus, the MLF treats all singly-occurring forms of different types and EL islands using the same

principles and procedures. As Myers-Scotton (2002) puts it, “[a] model that can cover all [EL] forms within the bilingual CP arguably is superior to one that cannot” (p. 153).

The opposition between ML and EL can be more formally stated in two principles:

(1) Key Principles of the MLF Model, restated from Myers-Scotton (1993 [1997])

a. *The Morpheme Order Principle*

In ML + EL constituents consisting of singly-occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.

b. *The (Late) System Morpheme Principle*<sup>1</sup>

In ML + EL constituents, all system morphemes which have grammatical relations external to their head constituents (i.e. which participate in the sentence’s thematic role grid) will come from the ML.

### 2.1.2 The content-system morpheme opposition

The second opposition, contrasting content and system morphemes, is key to understanding the connection between surface phrase structures and underlying abstract structures. Content morphemes can be considered the main elements of a construction. They carry the semantic/pragmatic information of an utterance. System morphemes serve mostly to show the relationships between content morphemes.

This opposition can be thought of as a split between frame-building properties. One such feature that sufficiently illustrates the distinction is [ $\pm$ thematic role assigner/receiver], or [ $\pm$ thematic role] for short. Although there are some syntactic categories, such as adjectives, for which the value of this feature is still debated, it is non-controversially accepted that nouns receive thematic roles, while most verbs (excluding the copula) and most prepositions assign thematic roles. For this work, we will analyze adjectives as content morphemes.

---

<sup>1</sup> This principle was originally called the System Morpheme Principle in Myers-Scotton (1993 [1997]). I add the “(Late) Morpheme” distinction because of later terminological developments from the 4-M Model (see section 2.4).

The content-system morpheme distinction, which implies that lexical elements are differentially accessed during language production, holds for both monolingual and bilingual speech. In code-switching however, this asymmetry clearly manifests itself in that content morphemes may be provided by either participating language, whereas system morphemes from the ML dominate.

## 2.2 Abstract Level Model

Unlike many other models, the MLF Model is lexically based. That is, rather than relying solely on principles of monolingual phrase structure to develop accounts for code-switching, the model underscores abstract procedures in and related to the mental lexicon. Some of these procedures involve phrase structure, but also include other abstract levels.

The Abstract Level Model, largely stemming from psycholinguistic models for language production (Levelt, 1993), was developed by Myers-Scotton and Jake (2000) as a supporting model to the MLF Model, but can also stand alone as a description of the levels of abstract lexical structure. The Abstract Level Model designates three levels of abstract lexical structure: lexical-conceptual structure, predicate-argument structure, and morphological realization patterns. A lemma, or an entry in the mental lexicon that maps abstract structure to surface realizations, is represented at all three levels. Thus, the Abstract Level Model serves to trace the path of a linguistic utterance from its beginnings as abstract structure to its manifestation as surface structure.

The origin of an utterance comes from an abstract bundle of language-independent speaker intentions. These intentions activate an abstract entity known as the Conceptualizer, which refines the message and decides what information is to be communicated linguistically and para-linguistically. The Conceptualizer triggers semantic/pragmatic feature bundles, and the ones that are language-specific are then mapped onto lemmas in the mental lexicon. This mapping forms the first level of abstract lexical structure, lexical-conceptual structure.

Once lemmas are active, their morphosyntactic properties (or instructions) can be accessed by the Formulator in order to generate hierarchical morphosyntactic structures. This requires two levels of structure, which involve the language-specific encoding or structural assignment of relations between content morphemes. The first of these two

levels, predicate-argument structure, deals with how thematic structure maps onto grammatical relations, and then morphological realization patterns deal with how grammatical relations map onto surface structures.

## 2.3 Congruence

In order for a form (i.e. a single morpheme or constituent) to appear in a given grammatical frame, a checking process must occur between that form and the specifications for its corresponding slot in the frame. Congruence can be understood as compatibility between a form (from any participating language) and its intended ML slot. In monolingual speech, this checking is trivial, because all forms are typically completely congruent with the slots they occupy.

However, complete congruence is rare, if even at all possible, in code-switching. In order for an EL form inserted into an ML frame to be acceptable, it must be checked for “sufficient congruence” with its frame. This checking takes place at the three levels of abstract grammatical structure.

The lack of sufficient congruence could potentially provide an explanation for impossibility of certain structures in code-switching, as well as for the types of compromise strategies that code-switching bilingual speakers sometimes use in order to compensate for insufficient congruence (Myers-Scotton, 1997).

The idea of congruence in bilingualism has been developed by Sebba (1998). According to this analysis, congruence in code-switching may not be inherently determined by the two participating languages alone, but rather, fueled by chosen and reinforced community norms. As stated by Sebba (1998), congruence is:

... not just a function of the syntax of the languages involved. The locus of congruence is the mind of the speaker, but community norms determine, by and large, the behavior of individual speakers. Bilinguals “create” congruent categories by finding common ground between the languages concerned. (p. 8)

## 2.4. The 4-Morpheme (4-M) Model<sup>2</sup>

The 4-M Model builds on the premise developed in the MLF Model that surface morpheme distributions can be best accounted for by considering how they are related to abstract structures in the mental lexicon. Specifically, the MLF Model expands on the content-system morpheme distinction, yielding four types of morphemes: content morphemes and three types of system morphemes known as early, bridge late, and outsider late system morphemes. One of the main aspects of this classification system is captured by the Differential Access Hypothesis:

(2) *The Differential Access Hypothesis*, restated from (Myers-Scotton and Jake, 2000)

The different types of morpheme under the 4-M Model are differently accessed in the abstract levels of the production process. Specifically, content morphemes and early system morphemes are accessed at the level of the mental lexicon, but late system morphemes do not become salient under the level of the Formulator.

This classification system was developed as a supporting model for the MLF Model, but it also stands alone as an independent model, empirically motivated by morpheme distributions in data sets from many fields of research, including on code-switching, interlanguage in second language acquisition, speech errors, and speech production by Broca's aphasics (Myers-Scotton, 2002).

In the paragraphs to follow, I will discuss this four-way morpheme distinction using aspects of both the original MLF Model and the 4-M Model that sprung from it.

---

<sup>2</sup> A note on terminology: Although true morphemes typically only refer to units of surface forms, in our discussion of the 4-M Model, we will use the term *morpheme* to refer collectively to a unit of surface structure, as well as its underlying lemma in the mental lexicon.

The four-way morpheme classification is a result of oppositions with regard to three features:

- (3) Oppositions in the 4-M Model (Myers-Scotton, 2002)
- a. [ $\pm$ conceptually activated] or [ $\pm$ conceptual]
  - b. [ $\pm$ thematic role assigner/receiver] or [ $\pm$ thematic role]
  - c. [ $\pm$ looks outside its immediate maximal projection for information about its form] or [ $\pm$ looks outside]

The first feature (3a) refers to the status of morphemes in abstract structure. If an element is conceptually activated, it is “salient as soon as the speaker’s intentions become encoded as language” (Schmid, 2004, p. 289), which is at the lemma level. The Conceptualizer maps speaker intentions onto language-specific semantic feature bundles, which we can consider the first level of linguistic structure. This feature distinguishes content and early system morphemes, which are both [+conceptual], from both types of late system morphemes.

The second feature (3b) was discussed previously in section 2.1.2. It distinguishes content morphemes, which are [+thematic role], from all system morphemes, but particularly from early system morphemes, with which content morphemes share the feature value [+conceptual].

Both types of late system morphemes are [-conceptual]. It is the third feature (3c) that differentiates outsider late system morphemes, which are [+looks outside], from all other morpheme types, but particularly from bridge late system morphemes, with which outsider late system morphemes share the feature value [-conceptual].

## 2.5. Summary

In section 2, we outlined an approach built out of several models that we will rely on throughout the rest of this paper. Now we summarize some of the main takeaways from section 2, as well as provide accompanying tables and figures to serve as references.

In section 2.1, we provided an overview of the MLF Model, which acknowledges and accounts for the asymmetry between the distributions of participating languages in code-switched speech. This asymmetry is the result of a key opposition between the ML and the EL (Table 1).

In section 2.2, we presented the Abstract Level Model, a supporting model to the MLF Model, which expands on the idea that the MLF Model is lexically based, unlike other accounts that rely on principles of monolingual phrase structure alone to account for code-switched speech. This discussion also allowed us to elaborate on a model of production that the MLF Model and its supporting models presuppose (Figure 1).

In section 2.3, we introduced the idea of sufficient congruence and its relevance in determining what code-switches are acceptable.

In section 2.4, we discussed the 4-M Model, which accounts for surface distributions of types of morphemes in a diverse range of phenomena. Three oppositions (Table 2) account for a four-way distinction (Table 3). Figure 1 provides a depiction of how these morpheme types are differentially accessed in the context of the presupposed model of language production.

<b>Language</b>	<b>Description</b>
Matrix Language (ML)	the language that contributes more abstract structure (i.e. language from which the grammatical frame of the CP is abstracted)
Embedded Language (EL)	the language that contributes less abstract structure (i.e. language inserted into the ML)

**Table 1**

The ML:EL opposition

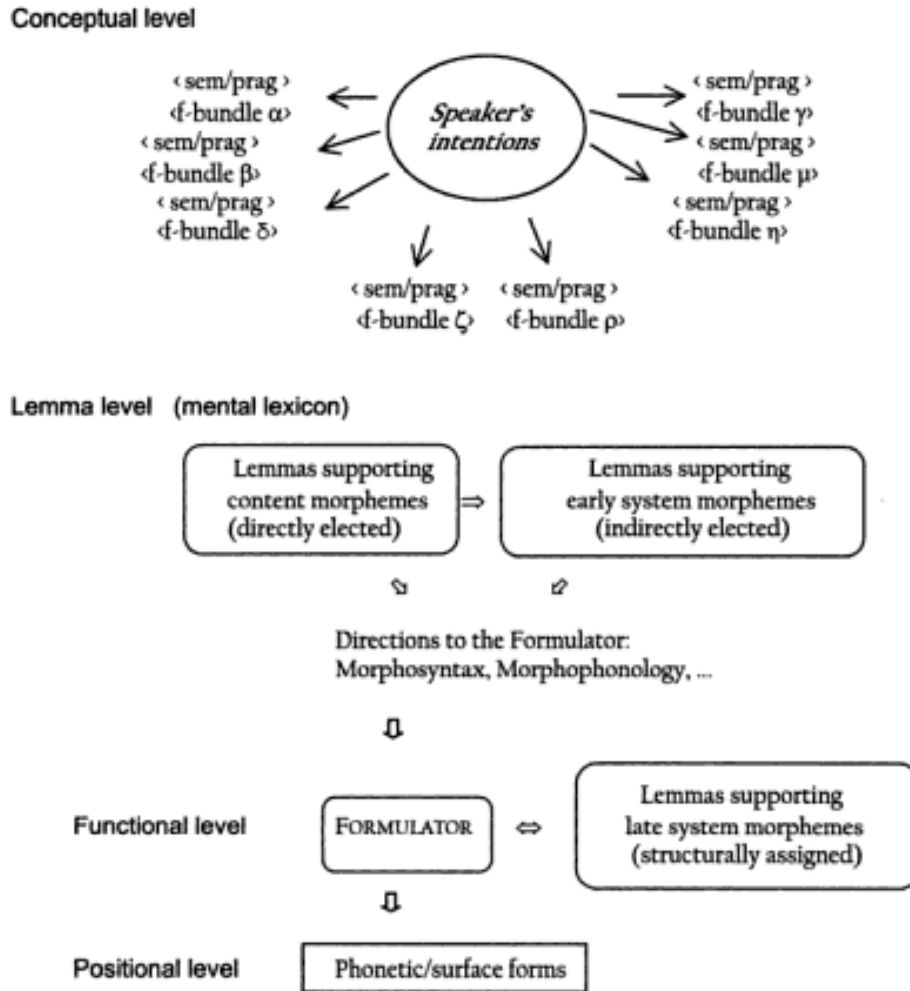


Feature	Feature Name, shortened	Description
conceptually activated	conceptual	<ul style="list-style-type: none"> <li>• [+] lemma salient at the level of the mental lexicon.</li> <li>• [-] lemma salient at the level of the Formulator.</li> </ul>
thematic role assigner/receiver	thematic role	<p>This opposition is relevant within [+conceptual]:</p> <ul style="list-style-type: none"> <li>• [+] lemma directly elected as head of maximal projection; contains information about well-formedness conditions for entire phrase projected by head.</li> <li>• [-] lemma indirectly elected by head; depends on head for information about its form.</li> </ul>
looks outside immediate maximal projection for information about its form	looks outside	<p>This opposition is relevant within [-conceptual]:</p> <ul style="list-style-type: none"> <li>• [+] lemmas activated when when large constituents (e.g. CP) are constructed; shows co-indexical relationships across maximal projections.</li> <li>• [-] lemma activated when grammatical configurations of a maximal projection requires it to complete the projection; shows co-indexical relationships within a maximal projection.</li> </ul>

**Table 2**  
Oppositions in the 4-M Model

Morpheme Type		Feature Values	Characteristic examples	
content		[+conceptual] [+thematic rol]	N, V, Adj	
system	early	[+conceptual] [-thematic role]	Spanish determiners	
	late	bridge	[-conceptual] [-looks outside]	English possessive <i>of</i> and <i>'s</i>
		outsider	[-conceptual] [+looks outside]	English subject-verb agreement, Latin case affixes

**Table 3**  
Morpheme types of the 4-M Model



**Figure 1**  
Production process diagram: lemma activation, from Myers-Scotton and Jake (2000)

## **3 Plural markers in Tagalog, English, and TECS**

In this section, we provide an introductory analysis of plural markers in monolingual Tagalog and monolingual English. We then use the approach presented in section 2 to account for the distribution of plural markers in TECS.

### **3.1 Plural markers in Tagalog and English**

In describing English and Tagalog plural markers, I adapt an analysis presented by Wiltschko (2008) used to compare plural markers in English and Halkomelem, a language spoken by about 600 in small indigenous communities in Southwest British Columbia (Lewis et al., 2013).

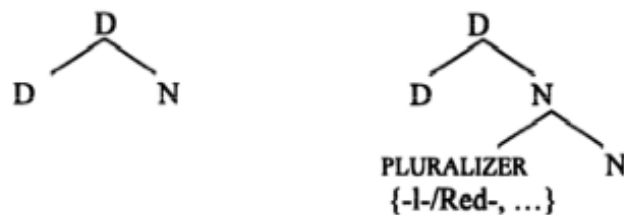
First, we start by considering nouns for both languages in the absence of plural markers. In English, an unmarked noun obligatorily has a singular reading. However, unmarked nouns in Tagalog are compatible with both the singular and the plural interpretation. In other words, Tagalog nouns lacking plural marker have properties associated with a “number-neutral interpretation” (Corbett, 2000) or “general number” (Rullman and You, 2006). I will refer to an element exhibiting such properties as “unspecified” for number. In effect, the result is that, on the surface, the Tagalog plural marker appears optional, since plural interpretations are possible whether or not the plural marker is present. This contrasts with plural interpretations in English, which require *-s*.

Wiltschko (2008) explains that this difference arises because of the syntactic nature of the markers. English *-s* is an instantiation of a value of the always present functional head Number, which has two possible values: singular and plural (Figure 2). Because of the status of the NumberP (denoted in Figure 2 by #) as an obligatory functional category, the singular value of this functional head must be analyzed as phonologically null.

**Figure 2**

The structure of the English noun phrase, from Wiltschko (2008)

However, data from Halkomelem calls into question the universality of the NumberP as a functional category. The analysis of Halkomelem by Wiltschko (2008) suggests that “plural marking is not universally merged as a syntactic (functional) head” (p. 639). Halkomelem provides one example, in which the Halkomelem plural marker acts as an adjoined modifier that “merges without the mediation of a functional category” (p. 646), i.e. it does not change the functional category of the constituent or element it merges with (Figure 3). Furthermore, there is no other value that contrasts with plural. The Tagalog plural marker *mga* has similar syntactic characteristics to those of the Halkomelem plural marker, as described in the analysis by Dionisio (2012) of *mga* as a modifier of one-place predicates (including nouns and nominal phrases). We will therefore be adopting a similar analysis of *mga* as an optional adjoined modifier.

**Figure 3**

The structure of the Halkomelem noun phrase, from Wiltschko (2008)

## 3.2 Plural markers in TECS

Considering the differences between Tagalog and English explored in section 3.1, the following question arises in the code-switching context:

- (4) How do code-switching bilingual Tagalog-English speakers deal with two differing strategies for nominal plural marker when inserting EL nominal phrases into the ML?

### 3.2.1 Observations

#### 3.2.1.1 *English as the ML*

When we insert a Tagalog N into an English frame (i.e. into an English N slot) we observe that *-s* is never allowed, whereas *mga* appears optional, just as in the monolingual Tagalog context. This is illustrated in (5):

- (5) (mga) bata(\*-s)  
'children'

#### 3.2.1.2 *Tagalog as the ML*

With Tagalog as the ML, we find a striking distribution of plural markers. It appears as though the English and Tagalog markers are both independently optional. Thus, all of the forms in (6) are compatible with a plural reading:

- (6) (mga) dog(-s)<sup>3,4</sup>  
'dogs'

Data illustrating this phenomenon have been observed in a corpus of utterances by Tagalog-English bilinguals compiled by Bautista (1980). Although she did not

---

<sup>3</sup> The form *mga dogs* is an instance of double morphology.

<sup>4</sup> The form *dog* may actually be glossed as (dog / dogs); that is, it can have both a singular or plural interpretation.

acknowledge the distribution of plural morphemes, let alone account for it, the corpus was full of examples, often with multiple tokens in a single speaking turn. Some noteworthy tokens of English EL nominal phrases with plural interpretation inserted into a Tagalog ML are provided in (7):

- (7) From Bautista (1980, p. 282-3)
- a. mga parke<sup>5</sup>
  - b. mga park
  - c. mga parks
  - d. national parks

### 3.2.2 Analysis

#### 3.2.2.1 *English as the ML*

These observations do not pose any issues for our analysis based on the MLF Model and its supporting models. The Tagalog EL N is a content morpheme that is simply inserted into the English frame. Election of the Tagalog N content morpheme does not trigger the indirect election of English early system morpheme *-s*. Thus, the lemma for *-s* never has a chance to be activated. Tagalog *mga*, on the other hand, may be analyzed as an adjective-like content morpheme. Due to its adjunct status, it is optional just as in monolingual Tagalog sentences. The status of *mga* as a content morpheme is supported by the fact that *mga* can serve as the only code-switched item in an English ML CP as in (8):

- (8) The boys ate all of the (*mga*) chocolate donuts.

PL

---

<sup>5</sup> The word *parke* ‘park’ can be analyzed as a Tagalog word, although probably originally a borrowing from the Spanish.

Thus, both the presence of the Tagalog EL N and the presence of *mga* can be analyzed as two separate simple insertions of Tagalog EL content morphemes into the English ML, for which sufficient congruence is achieved.

### 3.2.2.2 *Tagalog as the ML*

The distribution of plural markers in CPs that have Tagalog as the ML can be broken down into two phenomena. First, we analyze *mga* as functioning the same way in Tagalog ML sentences (i.e. either bilingual or monolingual sentences). That is, its behavior can be explained by its status as a content morpheme that is an optional adjoined modifier.

But how now do we account for the distribution of *-s* when Tagalog is the ML? We can state this question differently:

- (9) a. When the bilingual speaker wants to insert an English EL N into the Tagalog ML frame, why can the English *-s* plural marker appear at all?
- b. Why does *-s* appear optional, when in well-formed English phrases with plural interpretation, *-s* is obligatory?

The answers to these questions lie in the fact that inflectional plural affixes, such as English *-s*, are early system morphemes, i.e. “they add conceptual structure to their head and their form depends on their head” (Myers-Scotton, 2002, p. 92). This means that they can be accessed at the same level and at the same time as their content morpheme heads, the lemma level in the mental lexicon.

We can gain a better understanding of what happens by examining Myers-Scotton's (2002) treatment of double morphology, the most common example of which happens to be with plural affixes. Originally, Myers-Scotton (1993 [1997]) explained that double morphology was the result of some form of “mistiming.” The development of the 4-M Model provides us with the tools to characterize such a mistiming event. Both content and early system morphemes are [+conceptual]. Thus, such a mistiming is possible because, when a speaker’s intentions are mapped onto and activate a lemma for

an EL N, the lemma for EL nominal plurality is theoretically accessible and may also be activated.

A prediction based on the idea that such mistimings are possible is formalized as a hypothesis motivated by the 4-M Model in (10):

- (10) *The Early System Morpheme Hypothesis*, restated from Myers-Scotton (2002)  
Only early system morphemes may be doubled in classic code-switching.

The doubling of plural markers in TECS (e.g. *mga dogs* in (6)) does not directly support this hypothesis, because the Tagalog modifier *mga*, according to our analysis, is not an early system morpheme, but rather, a content morpheme. Thus, *mga dogs* from (6) does not technically illustrate an instance of early-system-morpheme doubling, as described in the Early System Morpheme Hypothesis (10).

From our data from TECS, we can make two observations (11) that are not accounted for by the Early System Morpheme Hypothesis (10):

- (11) a. There is an asymmetry between the *doubled* morpheme (i.e. the ML morpheme that is doubled) and the *doubling* morpheme (i.e. the EL morpheme that does the doubling).  
b. The *doubling* EL morpheme is not necessarily an early system morpheme (such as when English is the ML and we are inserting Tagalog *mga*).

In addition, I propose that the status of the doubled ML morpheme is not relevant to the actual event of interest, i.e. the activation of the doubling EL morpheme. This activation of the doubling EL morpheme takes place in lexical-conceptual structure and may either be direct activation (for content morphemes) or indirect activation (for early system morphemes). Thus, I present a revised version of the Early System Morpheme Hypothesis, which I am renaming the Double Morphology Hypothesis (12), that hones in on the events underlying the observations regarding double morphology in (6), which are summarized in (11):



(12) *The Double Morphology Hypothesis* (newly proposed in this work)

Only a [+conceptual] morpheme from the EL can double the meaning of an ML morpheme.

This revision accounts for the fact that the Tagalog EL content morpheme *mga* can double the plural meaning of *-s* in English ML CPs and the fact that the English EL early system morpheme *-s* can double the plural meaning of *mga* in Tagalog ML CPs.

The revised hypothesis essentially says that [+conceptual] morphemes can come from the EL, which in some cases results in double morphology. This claim is complementary in nature to the (Late) System Morpheme Principle, which states that [-conceptual] morphemes cannot come from the EL.

All the data that supported the original Early System Morpheme Hypothesis (10) should still support the revised version. However, now the revised hypothesis (12) acknowledges that lemmas are language-specific and similar meanings may be stored and encoded in different ways cross-linguistically.

Thus, we can offer answers to the questions asked in (9) that are consistent with the revised Double Morphology Hypothesis (12). The answer to question (9a) (regarding the possibility of inserting *-s*) is that early system morphemes like *-s* can be indirectly co-activated with their nominal heads (which are content morpheme Ns).

The answer to question (9b) (regarding the optionality of inserting *-s*) is that the mistiming that results in the surface presence of *-s* either may or may not happen. In terms of phrase structure, the speaker may be inserting the English EL N (which *is not* specified for number) into the Tagalog ML slot (as in (13b)), where it remains unspecified for number, effectively acting just like a Tagalog N (as in (13a)). Thus, (13b) might be an example of what some people consider a nonce borrowing, as it appears to be fully assimilated into the Tagalog morphosyntactic frame.

- (13) a. *Humanap ako ng kandila.*  
 searched for-AT<sup>6</sup> I NG candle  
 ‘I searched for (a candle / candles).’
- b. *Humanap ako ng candle.*  
 searched for-AT I NG  
 ‘I searched for (a candle / candles).’
- c. *Humanap ako ng child.*  
 searched for-AT I NG  
 ‘I searched for (a child / \*children).’
- d. *Humanap ako ng children.*  
 searched for-AT I NG  
 ‘I searched for (\*a child / children).’

The other possibility is that the speaker is inserting the English EL NumberP (which *is* specified for either singular or plural number). This is supported by the fact that the presence of *-s* on the English EL N requires the plural interpretation, as in (14).

- (14) (mga) dogs  
 ‘(\*dog / dogs)<sup>7</sup>

Further evidence that suggests the insertion of a NumberP is indeed possible comes from irregular nouns that require interpretation as already constructed NumberPs, as in (13c) and (13d).

Now, we can account for each of the forms in (6) separately. The form *dog* ‘(dog / dogs)’, whether it is considered a singly-occurring code-switch or an established borrowing (such as with cultural terms like ‘computer’ or ‘hamburger’), actually allows for both a singular interpretation and a plural interpretation. I propose that both interpretations are possible because the inserted form is actually the English EL N, unmarked and unspecified for number. In other contexts that don’t require a plural

<sup>6</sup> AT = Agent Topic. See section 4.1 for brief introduction to the Tagalog topic system.

<sup>7</sup> The notation ( A / B ) represents the idea ‘either A or B, but not both.’

interpretation, either the English EL N (unspecified for number) or the singular English EL NumberP, which on the surface appear identical, may be inserted. However, since there is no marker to distinguish the singular English NumberP from the English N (unspecified for number), the addressee may have to rely solely on context to figure out whether the speaker intended to produce a specifically singular interpretation.

For the form *mga dog* ‘dogs’, the inserted form is still an English EL N (unspecified for number), but the Tagalog ML content morpheme *mga* is present as an optional modifier. The inserted form cannot be analyzed as a singular English EL NumberP, because this would not agree with the meaning of *mga*.

For the form *dogs* ‘dogs’, the inserted form is an English EL NumberP specified for plural number and marked on the surface for this specification. The presence of the plural morpheme is due to the mistimed activation of the early system morpheme *-s*. A plural reading is obligatory.

For the form *mga dogs* ‘dogs’, in contrast with *dogs* ‘dogs’, simply reflects the optionality of the Tagalog ML modifier *mga* once more.

In sum, there is sufficient congruence between the English EL N and the Tagalog ML N slot, and there is also sufficient congruence between the English EL NumberP and the same Tagalog ML N slot.

### **3.2.3 Discussion**

#### ***3.2.3.1 Sufficient congruence is required for acceptable code-switches***

It may seem strange that we can insert English EL elements of differing functional categories, such as an N (i.e. an English noun unmarked and unspecified for number) and a NumberP (i.e. an English noun specified for number). However, to reiterate, it makes sense that this would be possible, so long as the requirement of sufficient congruence is obtained. We have analyzed Tagalog as not having a NumberP functional category, and thus, we can posit that code-switching Tagalog-English bilingual speakers deal with this difference between their languages by creating such congruent categories.

The English EL categories congruent with the slot associated with the Tagalog ML N do not stop there, as many types of English EL constituents (i.e. nominal phrases) can actually be inserted into this slot. (15) illustrates different possibilities for English EL insertion into a Tagalog ML N slot: N (15a), NP with modifier (15b) *and/or* NumberP with modifier (15c), NumberP + and + NumberP (15d), and a DP (15e).

- (15) a. (mga) [N]  
       mga topic (Bautista, 1980, p. 262)  
       ‘topics’
- b. (mga) [Adj + N]  
       mga wild animal (modified from (15c))  
       ‘wild animals’
- c. (mga) [Adj + NumberP]  
       mga wild animals (Bautista, 1980, p. 282)  
       ‘wild animals’
- d. (mga) [NumberP + and + NumberP]  
       mga [plants and animals] (Bautista, 1980, p. 283)  
       ‘plants and animals’
- e. (mga) [D + NumberP]  
       the facts of the matter (Bautista, 1980, p. 186)  
       ‘the facts of the matter’

One notable example in (15) is the instance of DP-insertion in (15e), since it can give yield to another instance of double morphology. Although the status of Tagalog nominal markers is still debated, they share properties with the English determiners and can be analyzed as determiners. This example (15e) might then be relevant for another case of double morphology resulting from the mistiming of the English EL early system morpheme *the*. Doubling of determiners has previously been observed in in the code-switching literature, including with TECS by Bautista (1980). (16) provides some tokens from Bautista (1980) of double determiners when Tagalog is the ML:

- (16) a. ng the facts of the case (p. 186)  
 NG  
 ‘the facts of the case.’
- b. ito pong the merits and demerits of the case (p. 186)  
 ANG-these RESP  
 ‘these merits and demerits of the case’

### 3.2.3.2 *Lemmas are language-specific*

Another interesting concept that our analysis underlines is that lemmas are language-specific. Even seemingly elemental meanings such as plural can be mapped differently in the mental lexicon depending on the language. In our example, it is important to keep in mind the cross-linguistic variation in nominal pluralization strategies, since the nature (with regard to the 4-M Model) of plural morphemes may differ among languages. Even Myers-Scotton (2002) does not account for this variation, treating determiners and ‘plural’ as characteristic and definitive examples of early system morphemes (p. 73). In our case, the English nominal plural marker *-s* is indeed an early system morpheme that includes morphosyntactic directions that mediate a functional category NumberP. The Tagalog nominal plural marker *mga*, on the other hand, is a content morpheme that includes morphosyntactic directions to generate a modified phrase by adjunction.

## 4 Subjecthood and case in Tagalog, English, and TECS

In this section, we provide an introductory analysis of the topic system and nominal markers in monolingual Tagalog. We also present an unresolved issue regarding subjecthood in Tagalog, as well as offer one possible explanation for this issue. We then use the approach presented in section 2 in an analysis of the distribution of Tagalog nominal markers in TECS.

## 4.1 Subjecthood and case in Tagalog and English

### 4.1.1 The Tagalog topic system

Each nominal phrase in a Tagalog sentence is marked with a default nominal marker, depending on what type of argument it is (i.e. Agent, Patient, Direction, and Benefactive). The status of these markers has been debated in the literature, having been analyzed and referred to as phrase-marker particles, construction markers, common noun markers, articles, determiners, specifiers, and proclitics (Reid, 2002).

In most Tagalog sentences, one of these nominal phrases has a special status as a topic. Philippine languages, such as Tagalog, are known for their complex topic systems. The basic structure of most Tagalog sentences exhibits apparent agreement between the topic and a particular morphology on the verb. This topic status is typically indicated by the nominal marker *ang* or one of its derivative forms (e.g. *iyong* and *yung*), all of which I will refer to as the *ang* marker. I will also refer to topics as *ang*-marked phrases.

Each type of argument can bear the *ang* marker, corresponding with different verbal morphologies depending on which argument is the *ang*-marked one; the result is a noteworthy paradigm of topic constructions that all seem to be equally morphologically marked. In Tagalog, the Agent, Patient, Direction, and Benefactive nominal phrases can each be *ang*-marked, corresponding with the four topic constructions similarly named (labeled AT, PT, DT, and BT, respectively, in (17)). When not marked with *ang*, an argument bears its default marker—*ng*, *ng*, *sa*, or *para sa*—based on whether it is the Agent, Patient, Direction, or Benefactive nominal phrase, respectively. Because *ang*, *ng*, *sa*, and *para sa* appear to serve a similar type of function in distinguishing between nominal phrases, I will refer to them all using the general term *nominal markers*. This avoids complications arising from the debate around the unresolved status of these morphemes as described above.

(17) The Tagalog topic system, from Schachter (1993)

a. Agent Topic construction

Sumulat ang bata ng liham sa abogado para sa babae.  
 wrote-AT ANG child letter lawyer woman.  
 ‘The child wrote a letter to a/the lawyer for a/the woman.’

b. Patient Topic construction

Sinulat ng bata ang liham sa abogado para sa babae.  
 wrote-PT child ANG letter lawyer woman  
 ‘A/the child wrote the letter to a/the lawyer for a/the woman.’

c. Directional Topic construction

Sinulatan ng bata ng liham ang abogado para sa babae.  
 wrote-DT child letter ANG lawyer woman  
 ‘A/the child wrote a/the letter to the lawyer for a/the woman.’

d. Benefactive Topic construction

Isinulat ng bata ng liham ang abogado ang babae.  
 wrote-BT child letter ANG lawyer woman  
 ‘A/the child wrote a/the letter to a/the lawyer for the woman.’

#### 4.1.1 The puzzle of subjecthood in Tagalog

There has been much debate about the role of the *ang*-marked phrase and the topic system, in general. In all the above examples (17a-d), the Tagalog sentences essentially have the same semantic meaning, except for the specificity requirement on the *ang*-marked phrase. However, there is also a nuanced, but distinctive notion of “aboutness” associated with the *ang*-marked phrase, which creates subtle differences in interpretation among the four topic constructions. Some people have thus equated the *ang*-marked phrase as a grammatical subject of sorts. While this suggests that *ang* is a subject marker, it is important to know that there are also competing analyses that have also analyzed *ang* as a nominative case marker, consistent with an accusative analysis of Tagalog (Kroeger, 1993), and an absolutive case marker, consistent with an ergative analysis of Tagalog (Aldridge, 2006).

Problems arise, however, when we compare grammatical subjects from other languages with that of Tagalog. There appears to be a split in grammatical properties normally associated with subjects (such as in languages in the Romance and Germanic families, including English) between two Tagalog nominal phrases (Table 4). Some of these properties always pattern with the *ang*-marked phrase, such as the various *ang*-marked nominal phrases in (17a-d). However, other properties normally associated with the subject actually pattern with the Agent phrase, such as (*ng / ang*) *bata* ‘a/the child’ (17a-d)), regardless of the particular topic construction. This presents a challenge to the notion of grammatical subject as we are familiar with it.

<b><i>ang</i>-marked nominal phrase</b>	<b>Agent nominal phrase</b>
1. Obligatory element of every clause 2. Launches floating quantifiers 3. Relativization	1. Reflexive binding 2. Equi target 3. Imperative addressee 4. Relevance to word order (based on other Philippine languages)

**Table 4**

The division of properties normally associated with subjects (as in English) between two nominal phrases in Tagalog (from Kroeger (1993))

The debate about the morphosyntactic system in Tagalog involves many competing and overlapping notions of subject, topic, voice, case, accusativity, and ergativity, etc. The different analyses of the *ang*-marked form deal with its determiner-like function, as well as its apparent function relating the argument it marks with the rest of the sentence. Without discussing or necessarily evaluating major competing views on these issues here, we adopt an analysis by Rackowski (2002), who analyzes *ang* as a subject marker, rather than a case marker, and provides a potential explanation for the split in properties described above.

Some analyses have referred to the *ang*-marked argument as being selected by the verbal morphology, while others yet analyze the verbal morphology as an instantiation of verbal agreement with the *ang*-marked nominal phrase. Rackowski (2002) instead argues that both the *ang* marker and the verbal morphology are reflexes of an Agree relation between the *ang*-marked phrase and T. Rackowski (2002) appeals to the Configurational Case Hypothesis, which posits that all cases in Tagalog are assigned in their base



positions. For example, in (17a-d), the Agent nominal phrase ‘a/the child’ has nominative case in all four topic constructions, while the *ang*-marked phrase, i.e. the subject, changes with each topic construction. This provides an explanation for the split in properties normally associated with subjects between the Tagalog Agent nominal phrase and the Tagalog *ang*-marked nominal phrase.

Thus, this analysis proposes that the set of properties we normally associate with grammatical subjects are actually an alignment of the set of properties associated with nominative case and the set of properties associated with grammatical subject. These entities are demonstrably decoupled in Tagalog.

## 4.2. Subjecthood and case in TECS

Considering the differences between Tagalog and English explored in section 4.1, the following question arises in the code-switching context:

- (18) How do code-switching Tagalog-English bilingual speakers deal with these differing strategies for subjecthood and case when inserting EL Ns/NPs/DPs into the ML?

### 4.2.1 Observations

#### 4.2.1.1 *Tagalog as the ML*

Limiting ourselves to examples with only two arguments (an Agent nominal phrase and a Patient nominal phrase), and thus, only two possible topic constructions, we find that inserted English nominal phrases cannot stand alone. That is, they must be marked with one of the Tagalog nominal markers, either *ang* or *ng*, depending on the requirements of the Tagalog CP. Thus, these two markers follow a similar distribution in code-switched CPs with Tagalog as the ML as they do in monolingual Tagalog.

#### 4.2.1.2 *English as the ML*

When English is the ML, the most prevalent insertion related to nominal phrases is the insertion of the *ang* nominal marker in place of what seems to be an English determiner or demonstrative like ‘the,’ ‘this,’ and ‘those.’

- (19) a. [Ang family planning component]... is the most crucial... (Bautista, p. 193)  
           ANG
- b. We’ll still discuss [ito                   pong mga details]. (Bautista, p. 194)  
   ANG-these RESP PL
- c. [Yung parents ] will discuss yung        decision.  
           ANG    ANG

We can analyze these insertions in another way. We can view them as insertions of mixed constituents consisting of a mostly English nominal phrase and a Tagalog nominal marker. It turns out that these mixed constituents behave almost identically to Tagalog EL islands, as illustrated by the modifications to (19) shown in (20).

- (20) Modified from (19)
- a. [Ang bahay]... is the most crucial...  
           ANG house
- b. We’ll still discuss [ito                   pong mga detalye].  
   ANG-these RESP PL details
- c. [Yung mga magulang] will discuss yung desisyon.  
           ANG PL    parents                                    ANG decision

Thus, we will include mixed nominal phrases under the designation of Tagalog nominal phrases, as long as the Tagalog nominal marker is present.

When we insert Tagalog nominal phrases into an English ML, it almost always bears the *ang* marker, regardless of the phrase’s status as English subject (nominative case) or object (accusative case). Specifically, we can make three observations:

- (21) a. Insertion of a Tagalog EL nominal phrase requires a Tagalog EL nominal marker, as in (19) and (20).
- b. This nominal marker is almost always *ang*, as in (19) and (20).
- c. There can be two *ang*-marked phrases in one CP (serving as English subject and object), as in (19c) and (20c).

These observations (21) pose an interesting puzzle for our analysis. We will refer to the phenomenon described by (21b), in combination with (21a) and (21c) as “the primacy of the *ang*-marked form”, from “the primacy of the *ang* form,” as Bautista (1980, p. 194) previously coined.

## 4.2.2 Analysis

### 4.2.2.1 *Tagalog as the ML*

English EL nominal phrases bear Tagalog nominal markers as if they were true Tagalog nominal phrases. Based on the analysis we have adopted from Rackowski (2002), we expect *ang* to behave as an outsider late morpheme. Although we will not delve into how exactly the other nominal markers are assigned, but we can similarly analyze them as morphemes that relate an argument to the rest of the CP. Based on this analysis and knowing they appear in complimentary distributions to *ang*, we can speculate that all the Tagalog nominal markers may also be outsider late system morphemes as well.

This does not pose any problems for our analysis based on the MLF Model. These data are consistent with the (Late) System Morpheme Principle (1b), since the ML is supplying these late system morphemes.

### 4.2.2.2 *English as the ML*

Based on the analysis of *ang* by Rackowski (2002), we can consider *ang* to be an outsider late system morpheme that plays a key role in the structure of the CP and must appear once, and only once, per CP. This brings up a few questions (22) that correspond with each of our observations in (21):

- (22) a. Why is it possible for Tagalog nominal markers to be inserted into the English ML, apparently defying the (Late) System Morpheme Principle (1b)?
- b. Why is this nominal marker almost always *ang*?
- c. Why can *ang* appear twice in one CP (in the subject and object positions)?

Question (22a) refers both to the insertion of a nominal marker as part of a Tagalog nominal phrase (20) and to the more striking, apparent insertion of the nominal marker alone (19). In section 4.2.1.2, we explained that the latter can be treated as a case of the former.

In Section 4.2.1, we saw that Tagalog nominal markers were always required for Tagalog nominal phrases, regardless of the ML. With Tagalog as the ML, this could be explained by considering these markers outsider late morphemes that related their respective arguments to the greater structure of the CP. However, this analysis is not consistent with the data for when English is the ML, as the (Late) System Morpheme Principle (1b) seems to be violated.

An alternative analysis may help us begin to reconcile why *ang* is able to be inserted into an English CP (question (22a)). Perhaps *ang* is a multimorphemic element, such as determiners in German, which have underlying lemmas for gender, number, and case. A multimorphemic element can be a form with multiple underlying lemmas, even though it might not have a neatly decomposable surface structure corresponding to those lemmas. For example, the Spanish determiner *la* consists of multiple morphemes for singular, feminine, and definite (although the singular meaning can be analyzed as coming from a null element contrasting with the plural *-s*).

Perhaps the reason why *ang* is able to be inserted into an English CP is that, in addition to consisting of an outsider late system morpheme, *ang* also consists of some sort of early system morpheme. This would be consistent with its previously noted determiner-like distribution and properties, and might be a way to reduce the degree of the apparent violation of the (Late) System Morpheme Principle (1b).

However, this analysis remains problematic because it conflicts with a hypothesis from Myers-Scotton and Jake (2000), which was made based on data analyzed using the

MLF Model and stems out of the (Late) System Morpheme Principle (1b). The hypothesis is stated as Hypothesis 6 in Myers-Scotton (2002):

(23) *Hypothesis 6*, restated from Myers-Scotton (2002)

In multimorphemic elements (consisting of two or more system morphemes and including a late system morpheme), the late system morpheme takes precedence. This means that the entire element shows distribution patterns as if it were a late system morpheme. This is the ‘pull down’ or ‘drag down’ principle.

At this point in our analysis, we can either take our data to be counterevidence to this hypothesis (23) or evidence against our analysis of the *ang* marker based on Rackowski (2002).

Another possible analysis could be that *ang* is not a relational marker or subject marker at all. Instead, *ang* might be only an early system morpheme that marks specificity. Rackowski (2002) argues that both the *ang* marker and verbal morphology are reflexes of an Agree operation that involves looking outside the immediate maximal projection of the nominal phrase (i.e. they are both outsiders late system morphemes). If *ang* is indeed actually only an early system morpheme, this could mean that verbal morphology is simply a form of agreement with the *ang*-marked argument. This agreement may require overt or covert movement, but such details will not be explored here. In other words, verbal agreement would still be an outsider late morpheme, but *ang* would be an early system morpheme. This analysis would mean that Rackowski’s (2002) analysis based on the Configurational Case Hypothesis is incorrect.

However, we do not have enough information about the characterization of *ang* in order to determine whether our data should be interpreted as counterevidence to (24a) or (24b):

- (24) a. The pull down principle (Hypothesis 6 (22))  
 b. Rackowski’s (2002) analysis about case and subjecthood in Tagalog



## (26) Ditransitive verbs in active voice constructions

- a. (The/ yung / \*noong ) children gave (the/yung / \*noong) adults  
 ANG-those NG-those ANG-those NG-those

(the/ yung / noong ) details.

ANG-those NG-those

‘The children gave the adults the details.’

- b. (The/yung / \*noong ) children gave (the/yung / noong ) details to  
 ANG-those NG-those ANG-those NG-those

(the/ yung / \*noong ) adults.

ANG-those NG-those

‘The children gave the details to the adults.’

## (27) Ditransitive verbs in passive voice constructions

- a. (The/ yung / \*noong ) adults were given (the/ yung / noong )  
 ANG-those NG-those ANG-those NG-those

details.

‘The adults were given the details.’

- b. (The/ Yung / \*Noong ) details were given to (the/ yung / \*noong )  
 ANG-those NG-those ANG-those NG-those

adults.

‘The details were given to the adults.’

The above data in (26) and (27) show that within multiple active and passive voice constructions, the *ng* marker (in addition to the *ang*-marked form) is acceptable only on the direct object when it bears accusative case (indicated by underlining). At present, we are not able to propose an analysis for this data. Further work will be necessary to account for both the acceptability of the *ng*-marked form and its apparent interchangeability with the *ang*-marked form.

## 5 Conclusion

### 5.1 Summary

This study indeed proved the value in pursuing the Tagalog-English pair in code-switching research. The two languages are typologically very different and present interesting questions about how bilingual speakers deal with differing linguistics strategies. This research also shows that the MLF Model is a useful framework for analyzing TECS.

In section 1, we briefly introduced the concept of code-switching and Tagalog-English bilingualism, specifically the practice of TECS. We also provided some opening remarks in order to set the stage for the rest of the paper.

In section 2, we provided a theoretical framework that we would use for analysis of TECS. The core of this framework is the MLF Model. We also delved into supporting models and concepts: the Abstract Level Model, congruence, and the 4-M Model.

Section 3 was an investigation of issues of internal syntax in the nominal domain. We started with a characterization of nominal plural markers in monolingual Tagalog and monolingual English based of off ideas by Dionisio (2012) and Wiltschko (2008). Then we presented data from TECS, with both Tagalog and English as the ML. The 4-M Model proved particularly powerful as a tool to explain the observed distributions of plural markers. The data also served to highlight the concept of sufficient congruence and the language-specific nature of lemmas in the mental lexicon.

Section 4 was an investigation of issues of external syntax in the nominal domain. We opened by explaining the topic system of Tagalog and its nominal marker system. We presented a challenging issue regarding an apparent split in subject properties between two nominal phrases in Tagalog, and offered an account by Rackowski (2002) in order to explain this puzzle. We then used our theoretical framework in order to perform our analysis, in both Tagalog ML and English ML contexts. We found that, depending on our analysis of the *ang*-marked form, our data may either contradict the pull-down



principle (22) hypothesized by Myers-Scotton (2002) or the analysis based on the Configurational Case Hypothesis from Rackowski (2002).

## 5.2 Future directions

There remains much to be done in the field of code-switching. The MLF Model has become a very important model in analyzing bilingual speech, as well as many other types of contact phenomena. The fact that the empirical motivations of the MLF Model and its supporting models come from many different lines of research and the fact that these models have tremendous explanatory power in explaining many different types of data suggest that the MLF has been a big step in the right direction. This framework has so far proven useful not only when studying code-switching, but also when thinking about linguistic competence and performance more generally. However, these models also rely on concepts that need further testing and characterization in order to further validate these models and uncover the precise mechanisms of how the bilingual brain works. For example, the idea of congruence is a very important concept in mediating between a bilingual speaker's two languages. However, it certainly needs further development. Since it seems to have both structural and sociolinguistic components to it, but research from both structural and sociolinguistic approaches can contribute to our knowledge of congruence.

Further work is also needed on the characterization of Tagalog morphosyntax in general. The nature of the topic system and the status of *ang* and the other nominal markers are still debated issues. Syntactic, semantic, historical, and cross-linguistic analyses have all contributed to the effort of solving this mystery, but an undisputed analysis has yet to emerge.

Lastly, the addition of more data from previously studied language pairs, and of course, data from language pairs not represented in the code-switching literature will serve an important role in refining and differentiating between competing models of code-switching.

## References

- Aldridge, E. (2006). Absolute case in Tagalog. In Proceedings from the Annual Meeting of the Chicago Linguistic Society, pp. 1–15.
- Amuda, A.A. (1994). Yoruba/English Conversational Code-Switching as a Conversational Strategy. *Afr. Lang. Cult.* 7, 121–131.
- Bautista, M.L.S. (1980). The Filipino bilingual's competence: a model based on an analysis of Tagalog-English code switching (Canberra, A.C.T., Australia: Dept. of Linguistics, Research School of Pacific Studies, Australian National University).
- Corbett, G.G. (2000). *Number* (Cambridge University Press).
- Dionisio, M. (2012). *The Syntax and Semantics of the Tagalog Plural Marker Mga*. The Ohio State University.
- Finlayson, R., Calteaux, K., and Myers-Scotton, C. (1998). Orderly Mixing and Accommodation in South African Codeswitching. *J. Socioling.* 2, 395–420.
- Jake, J.L., Myers-Scotton, C., and Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Biling. Lang. Cogn.* 5.
- Kroeger, P. (1993). *Phrase Structure and Grammatical Relations in Tagalog* (Center for the Study of Language (CSLI)).
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.), (2013). *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Levelt, W.J.M. (1993). *Speaking: From Intention to Articulation* (MIT Press).
- Myers Scotton, C. (1983). The negotiation of identities in conversation: a theory of markedness and code choice. *Int. J. Sociol. Lang.* 1983, 115–136.
- Myers-Scotton, C. (1997). *Duelling Languages: Grammatical Structure in Codeswitching* (Oxford University Press, USA).
- Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes* (Oxford University Press, USA).
- Myers-Scotton, C., and Jake, J. (2000). Four types of morpheme: evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38.
- Rackowski, A. (2002). Subject and specificity: the case of Tagalog. In PROCEEDINGS-NELS, pp. 471–486.

Reid, L.A. (2002). Determiners, Nouns, or What? Problems in the Analysis of Some Commonly Occurring Forms in Philippine Languages. *Ocean. Linguist.* 41, 295–309.

Rullman, H., and You, A. (2006). General Number and the Semantics and Pragmatics of Indefinite Bare Nouns in Mandarin Chinese. In *Where Semantics Meets Pragmatics*, K. von Stechow, and K.P. Turner, eds. (Amsterdam: Elsevier), pp. 175–196.

Schachter, P. (1993). *The Subject in Tagalog: Still None of the Above* (University of California, Los Angeles.).

Shin, Hyon B. and Robert A. Kominski. (2010). *Language Use in the United States: 2007, American Community Survey Reports, ACS-12*. U.S. Census Bureau, Washington, DC.

Schmid, M.S. (2004). *First language attrition: interdisciplinary perspectives on methodological issues* (John Benjamins Publishing).

Sebba, M. (1998). A Congruence Approach to the Syntax of Codeswitching. *Int. J. Biling.* 2, 1–19.

Wiltschko, M. (2008). The Syntax of Non-Inflectional Plural Marking. *Nat. Lang. Linguist. Theory* 26, 639–694.