

Codeswitching in the
Multilingual Mind

by

Dustin Hilderman
BA, Simon Fraser University, 2012

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

© Dustin Hilderman, 2017
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

Supervisory Committee

Codeswitching in the
Multilingual Mind

by

Dustin Hilderman
BA, Simon Fraser University, 2012

Supervisory Committee

Dr. John Archibald, (Department of Linguistics)
Supervisor

Dr. Martha McGinnis-Archibald (Department of Linguistics)
Departmental Member

Dr. Suzanne Urbanczyk, (Department of Linguistics)
Departmental Member

Abstract

Supervisory Committee

Dr. John Archibald, (Department of Linguistics)

Supervisor

Dr. Martha McGinnis-Archibald (Department of Linguistics)

Departmental Member

Dr. Suzanne Urbanczyk, (Department of Linguistics)

Departmental Member

The very existence of intra-word codeswitching—of the type $[w M_{L1} + M_{L2}]$; $*[eat]_{eng} + [-iendo]_{spanish}$ —has long been a point of contention in the language mixing literature (Poplack, 1980; Myers-Scotton, 1992; MacSwan, 2005). However, recent work by Alexiadou et al (2015) and Grimstad, Lohndal & Afarli (2014) has documented a number of empirical examples of such codeswitching in an American community of Heritage Norwegian-English speakers—crucially, in these examples, the lexical elements are English lexical roots and produced using English phonological rules but the suffix (i.e. morphology) attached to the lexical items is syntactically Norwegian—a clear and unambiguous example of intra-word codeswitching. These data will be the focus of investigation into intra-word codeswitching.

MacSwan (2005) has argued that intra-word codeswitching is prohibited due to the inability of the human computational system to merge hierarchically ordered phonological systems from two or more languages; a prohibition characterized in his PF Disjunction Theorem. More recently, Alexiadou et al., (2015); Grimstad, Lohndal & Afarli, (2014) have exploited a model of Distributed Morphology to challenge the PF disjunction theorem and the ban on intra-word codeswitching it entails. A central goal of this thesis will be to compare, contrast and evaluate these two models of language mixing. It will be argued that this prohibition of intra-word language mixing may be overcome by appealing to a cognitive processes perspective (Sharwood-Smith & Truscott, 2014).

A MOGUL processing prospective (Sharwood-Smith & Truscott, 2014) will be used to build upon previous approaches to language mixing in order to account for intra-word codeswitching. The modular architecture adopted by MOGUL allows for a molecular view of a lexical item; each module (i.e. phonological module, syntax module,

conceptual module) produces a representation for a given form which is then interfaced to neighboring modules; the result is a chain of representations (i.e. PS + SS + CS) which constitutes a lexical item. Additionally, MOGUL incorporates several extra-linguistic cognitive mechanisms which play a role in language mixing. Of particular interest are the notions of goals and cognitive context. Following Sharwood-smith & Truscott (2016), goals are the central motivators for speech and action while cognitive context is taken to be the mentally internalized representation of an individual's current environment (Sharwood-Smith & Truscott, 2014) as well as representing various intentions, perspectives, opinions, etc., an individual has regarding their environment (Van Dijk, 1997).

To situate intra-word codeswitching into a MOGUL framework, much of MacSwan's Minimalist account will be adopted, (i.e. codeswitching is accounted for via the union of grammar X and grammar Y; formally: $\{G_x \cup G_y\}$) while rejecting the PF Disjunction Theorem and, instead, adopting elements of Distributed Morphology (i.e. late insertion). It will be argued that cognitive context configures various executive control process (i.e. bilingual mode) to allow for the union of phonological systems between L_x and L_y . This analysis builds upon a larger body of language mixing research by synthesizing a Minimalist account of codeswitching with a cognitive processing framework to account for intra-word codeswitching; the MOGUL framework allows for these disparate elements to be synthesized.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Abbreviations	ix
Chapter 1: Introduction	1
1.0.0. Introduction: The Puzzle.....	1
1.0.1. Introduction: A MOGUL Solution.....	3
1.0.2. Thesis Overview	3
1.1.0. MOGUL.....	5
1.1.1. Why MOGUL?	6
1.1.2. What is MOGUL?.....	7
1.1.3. Advantages of an Interdisciplinary Framework.....	8
1.2.0. Theoretical Challenges for a Cognitive Framework of Language.....	9
1.3.0. Chapter Summary	11
Chapter 2: The Cognitive Framework	13
2.0.0. Cognitive Framework	13
2.1.0. MOGUL Architecture.....	13
2.1.1. Basic Architecture.....	14
2.1.2. Inter-modular Relations: General	15
2.1.3. Inter-modular Relations: Key Modules	16
2.1.4. Inter-modular relations: PS < - > SS (The Language Core)	16
2.1.5. Inter-modular relations: SS < - > CS	17
2.1.6. Inter-modular relations: AS < - > PS	18
2.1.7. Inter-modular relations: AS < - > CS.....	19
2.1.8. Inter-modular relations: VS < - > PS	19
2.1.9. Inter-modular relations: CS < - > POpS & AffS	20
2.2.0. MOGUL and Modularity	22
2.2.1. Information Stores & Representations.....	23
2.2.2. Integrated Processors & Active Content.....	24
2.2.3. Interface Processors, Co-indexation & Chains	26
2.2.4. Conceptual Representations, Metalinguistic Knowledge & Lexical Meaning... ..	28
2.2.4. Lexical Entries, the Lexicon & Memory	30
2.2.5. Lexical Access: Dual Storage & The Processing Race	31
2.2.6. Functional vs. Lexical Categories.....	33
2.2.7. The Nature of Processing in MOGUL.....	33
2.3.0. Acquisition by Processing Theory (APT) & States of Activation.....	34
2.4.0. MOGUL & Cognitive Context	37
2.4.1. Cognitive Context & Real-time Processing.....	39
2.4.2. Influences on Cognitive Context	39
2.4.3. Cognitive Context: An Example.....	41
2.4.4. How is Context Realized in MOGUL.....	43

2.4.5. Cognitive Context, Conceptual Triggering & Language Selection	44
2.4.6. Evidence Supporting Cognitive Context	47
2.5.0. MOGUL & Executive Control.....	49
2.5.1. GOAL Representation in MOGUL.....	50
2.5.2. MOGUL and Communicative Modes.....	52
2.6.0. Chapter Summary	53
Chapter 3: The Linguistic Framework	55
3.0.0. Language Mixing	55
3.1.0. Codeswitching & Borrowing	56
3.1.1. Codeswitching and Speech Mode	58
3.2.0: Generative Grammar & Language Mixing	61
3.2.1: Constraints on Codeswitching	61
3.2.2: MacSwan, MP and the Language Faculty	63
3.2.3. Intra-word Codeswitching and the PF Disjunction Theorem	65
3.3.0. A Brief review of Distributed Morphology and Language Mixing	68
3.3.1. Distributed Morphology: The Syntactic Frame	68
3.3.2. Distributed Morphology: The Lists	70
3.3.3. DM & Language mixing	73
3.3.4. Norwegian-English Language Mixing.....	76
3.4.0. Chapter Summary & Conclusion	79
Chapter 4: A Unified Framework	81
4.0.0. MOGUL and Intra-word Language Mixing:	81
4.1.0. The Cognitive Framework: Review	81
4.1.1. The Linguistic Framework: Review	84
4.2.0 Situating DM in a MOGUL Framework: Bringing it all together	86
4.2.1 Intra-word Codeswitching in MOGUL: An Empirical Example.....	91
4.3.0 Discussion: MOGUL Predictions on Intra-word Codeswitching	97
4.3.1 Discussion: From Codeswitching to Loanwords	100
Section 4.4.0 Chapter Summary:	102
Chapter 5: Conclusion.....	104
5.0.0. The Future of MOGUL.....	104
5.0.1. Future Directions: The Quantum Metaphor.....	106
5.1.0 Final Summary & Conclusion	107
Bibliography	110

List of Tables

Table 1: Types of Codeswitching	59
Table 2: Review of Proposed Constraints on Codeswitching.....	62
Table 3: Norwegian Nominal Affixes with Examples.....	76
Table 4: English Loanwords in Mexican Spanish	102

List of Figures

Figure 1: Tripartite Model	14
Figure 2: Key Relation between Modules	19
Figure 3: MOGUL Architecture	21
Figure 4: Diagram of a Module in MOGUL.....	23
Figure 5: Diagram for ‘Pat hit Chris’	27
Figure 6: Conceptual Representation for Lamp.....	29
Figure 7: The Molecular Metaphor: A word as an atom/molecule.....	31
Figure 8: Learning in APT	35
Figure 9: Street signs contribute to a language user’s linguistic landscape.....	40
Figure 10: Multilingual street signs contribute to a multilingual linguistic landscape.....	40
Figure 11: General Language Representations (GLRs).....	46
Figure 12: Stimuli from Blanco-Elorrieta & Pylkkanen (2015: p.2).....	48
Figure 13: A Bilingual Minimalist Grammar	65
Figure 14: Norwegian Syntactic Frame	69
Figure 15: Stages of the Derivation	71
Figure 16: A DM Model of a Bilingual Grammar.....	75
Figure 17: Syntactic Structure for den field-a.....	78
Figure 18: Feature Co-indexation	93
Figure 19: Feature Co-indexation with Oscillating Contexts	94
Figure 20: Feature Co-indexation with Oscillating Context and a Goal	95
Figure 21: Derivation for Den Fielda.....	96

List of Abbreviations

General Abbreviations:

DM – Distributed Morphology
 UG – Universal Grammar
 MP – The Minimalist Program
 LAD – Language Acquisition Device
 LTM – Long Term Memory
 ACC – Anterior Cingulate Cortex

MOGUL Abbreviations:

MOGUL – Modular Online Growth and Use of Language
 PS – Phonological Structures
 SS – Syntactic Structures
 CS – Conceptual Structures
 POPs – Perceptual Output Structures
 AS – Audio Structures
 VS – Visual Structures
 GS – Gastronomical Structures
 OS – Olfactory Structures
 TS – Tactile Structures
 KS – Kinesthetic Structures
 AffS – Affective Structures
 APT – Acquisition by Processing Theory
 GLR – General Language Representation

Syntax Abbreviations:

1S – First Person Singular
 3S – Third Person Singular
 DUR – Continuous
 FUT – Future Tense
 PAST – Past Tense
 DEF – Definitiveness
 Gen – Gender
 F – Feminine
 M – Masculine
 Num - Number
 D – Determiner
 MS – Memory Store
 Lex – Lexicon
 C_{HL} – Computational System for Human Language
 PF – Phonological Form
 LF – Logical Form

Chapter 1: Introduction

1.0.0. Introduction: The Puzzle

Language mixing refers to a language user's ability to combine elements from two separate languages (or language varieties) in speech. While the phenomenon has been well documented over the years, there are a number of questions which remain unanswered: how can a speaker combine two language systems? Are languages stored separately in a bilingual brain? Are there special mechanisms or processes in the brain that allow languages to be mixed or is language mixing a natural product of language processing? These questions have attracted researchers from a variety of disciplines including generative linguists (Chomsky, 1995; Jackendoff, 2003; MacSwan, 2014), cognitive scientists (Carroll, 2001; Sharwood-Smith & Truscott, 2014), sociolinguists (Poplack, 1980; Poplack, Sankoff and Miller, 1988) and philosophers (Fodor, 1983); all attempting to solve their own little piece of the language mixing puzzle.

While there is a plethora of mysterious and fascinating linguistic phenomena observed in language mixing, this thesis is particularly interested in a specific type of language mixing known as intra-word codeswitching—that is to say, a speaker's ability to combine elements from two languages within a single word. Historically, a number of prominent codeswitching researchers have argued that intra-word codeswitching, where a word from one language is used as the base of a morphologically complex form in the other language, should be impossible (see Poplack, 1980; Poplack, Sankoff & Miller, 1988; MacSwan, 2005, 2014). Poplack (1980) provides the following examples of ungrammatical intra-word codeswitching.

- 1a) *Juan esta eat-iendo
Juan be/1Ss eat-DUR
Juan is eating.

- 1b) *Juan eat-o
Juan eat-PAST/3Ss
Juan ate.

- 1c) *Juan eat-ara
 Juan be/1Ss eat-FUT/3Ss
 Juan will eat.

(Poplack 1980)

In colloquial terms, MacSwan (2005) argues that language mixing at the word level (i.e. intra-word codeswitching) should be impossible as each language has its own set hierarchically ordered phonological rules (Bromberger & Halle, 1989) which cannot be combined in a cogent way (see chapter 3 for a more technical discussion). This thesis will contest this claim. Recent work by Alexiadou et al. (2015) and Grimstad, Lohndal & Afarli (2014) has documented a number of empirical examples of this type of codeswitching in an American community of Heritage Norwegian-English speakers. This heritage community largely immigrated in the latter half of the 19th century and settled in the upper-midwestern United States (e.g. Minnesota area). Notably, this community has continued to thrive over the years and have maintained their cultural and linguistic heritage—there are roughly 40,000 – 50,000 bilingual speakers of Heritage Norwegian currently living in the U.S. according to a 2000 U.S. census; while English is the language of the law and used at work and at school, Norwegian is still commonly spoken in domestic situations. These data, provided by Alexiadou et al. (2015), will be the focus of the present investigation into intra-word codeswitching.

- 2a) den **field**-a
 that field-DEF.F
 ‘that field’

- 2b) den **track**-en
 that track-DEF.M
 ‘that track’

(Alexiadou et al. 2015)

Crucially, in these examples, the elements *field* and *track* are English lexical items and produced using English phonological rules, but the suffixes attached to the lexical items

are Norwegian—a clear and unambiguous example of intra-word codeswitching. These data demonstrate, unequivocally, that the phenomenon of intra-word codeswitching is, in fact, very real. So, the puzzle remains, in these instances of word-level language mixing, how is it possible for two separate phonological systems to contribute in the formation of a single word (i.e. an intra-word codeswitch)? This is the central question which will be explored throughout this thesis.

1.0.1. Introduction: A MOGUL Solution

This thesis is interested in the production of intra-word codeswitches and, as such, will adopt a performance-based framework of language known as the **Modular Online Growth and Use of Language (MOGUL)** framework (Sharwood-Smith & Truscott, 2014 - hereafter SS&T). MOGUL offers a holistic perspective of the multilingual mind/brain which incorporates extra-linguistic operations (e.g. sensory perception, emotion, context, etc.) into a model of language performance in real-time. More precisely, this framework is primarily interested in the modular representation and processing of language—the MOGUL framework will be explicated in detail in chapter 2. It will be a central goal of this thesis to argue that intra-word codeswitching is a natural product of language processing in the MOGUL framework. Of interest to this research is the nature of the cognitive mechanisms that allows for intra-word codeswitching to occur. Specifically, this thesis seeks to understand the machinery in the mind/brain which allows for the mixing of multiple phonological systems within a word.

1.0.2. Thesis Overview

This thesis seeks to make a novel contribution to the topic of codeswitching by coalescing diverse research on this topic from a variety of academic disciplines. This will be accomplished by using the phenomenon of codeswitching to form a bridge between generative grammar and psycholinguistics which will be exploited to develop a performance-based account of codeswitching; more specifically, the MOGUL framework (Sharwood-Smith & Truscott, 2014)—a modular, performance-based framework of the multilingual mind—will provide the cognitive architecture necessary to account for word-internal (i.e. intra-word) codeswitching.

To explore language-mixing in the mind/brain, this thesis will review English-Norwegian language mixing data from Grimstad, Lohndal and Afarli (2014) and Alexaidou et al. (2015). Two Minimalist accounts to language mixing will be investigated: (1) a Lexicalist approach presented in MacSwan (2005, 2014); (2) a Distributed Morphology approach advocated by Alexaidou et al (2015) and Grimstad, Lohndal and Afarli (2014). Notably, these generative approaches to language mixing are competence-based models of representing language which focus on language mixing in I-language¹.

This thesis will aim to provide a cognitive account of intra-word codeswitching in the mind/brain. To do so, I will attempt to situate a Minimalist account of codeswitching into the MOGUL framework (SS&T, 2014); a performance-based model of language which is interested in the real-time representation and processing of language. While this thesis adopts major aspects of MacSwan's model (i.e. the grammatically unconstrained nature of code-switching; see section 3.2.0 below for details), there are concerns with his model. Most notably, this thesis will reject MacSwan's PF Disjunction Theorem and the prohibition on intra-word codeswitching it entails—a claim born of MacSwan's lexicalist assumptions. Central to this position, this thesis will argue that cognitive context (SS&T, 2014; Van Dijk, 1997) and bilingual mode of communication will allow for the union of multiple phonological systems which, in turn, allows for the mixing of phonological systems during intra-word codeswitching.

Additionally, this thesis will challenge MacSwan's lexicalist assumptions in favor of a Distributed Morphology perspective (Halle & Marantz, 1994); notably, this thesis will adopt a Late-Insertion Model where morpho-phonological elements are inserted into the derivation during PF-Spellout by the operation Vocabulary Insertion (Alexaidou et al. 2015). This perspective follows recent work by Alexaidou et al. (2015), Gonzalez-Vilbazo & Lopez (2011) and Grimstad, Lohndal & Afarli (2014) who appeal to a Late-Insertion Model (i.e. Last-Resort Model² in Gonzalez-Vilbazo & Lopez) of Distributed Morphology (DM) to account for intra-word codeswitching data. This approach will then

¹ I-language is the internal mental structure of Language; I-language is Individual, Internal & Intentional.

² While a 'Late-insertion model' is not synonymous with a 'Last-resort Model', it is my understanding that both models rely on Vocabulary items being inserted into a fully constructed syntactic frame.

be adapted to the MOGUL context with the goal of investigating how elements involved in language switching are represented and processed in the mind/brain. Ultimately, this thesis attempts to build an account of language mixing which situates a DM model of linguistic competence in a MOGUL model of linguistic performance.

Chapter 1 proceeds with a brief introduction to MOGUL and a theoretical justification as to why MOGUL is a suitable framework for this investigation into intra-word codeswitching. The motivations behind the MOGUL framework will be highlighted and the advantages of such an interdisciplinary model will be discussed. Chapter 2 will focus on the cognitive framework; the MOGUL architecture will be introduced and reviewed in detail. Additionally, I will discuss key cognitive mechanisms and operations that will be relevant to my account of intra-word codeswitching—namely, cognitive context and executive control. Chapter 3 will discuss the linguistic framework and will begin by providing a brief review of current approaches to codeswitching. Special attention will be given to MacSwan’s lexicalist approach to codeswitching, and its implications for intra-word codeswitching. Following this, I will present work from Grimstad, Lohndal & Afarli (2014) who exploit elements of DM to account for intra-word codeswitching. Chapter 4 will bring together chapters 2 and 3 into a MOGUL account of intra-word codeswitching by explicating in detail how elements of DM fit into a MOGUL framework. Consideration will be given to the advantages of such an approach. Chapter 5 will conclude with a brief summary and a discussion of future research directions.

1.1.0. MOGUL

MOGUL (Modular Online Growth and Use of Language) is a modular perspective on language processing presented by Sharwood-Smith & Truscott (SS&T, 2014), with the goal of explaining how language inhabits the mind in real time. A particular focus is placed on modularity (Fodor 1983), including how language is represented in each module and how these modules interact with each other. Ultimately, SS&T set out to design a collaborative framework in which various psychological, neurological and linguistic theories can be incorporated into a more holistic view of language and language processing in the mind/brain. As such, this makes MOGUL an ideal cognitive framework in which to compare competing linguistic theories of the

phenomenon of language mixing (Dijkstra & Haverkort, 2004; Roeper, 2004; Sharwood-Smith & Truscott, 2004).

1.1.1. Why MOGUL?

Before we delve into the MOGUL framework, it may be prudent to offer a justification as to why the MOGUL framework is a suitable platform for the present exploration into intra-word codeswitching. The first major advantage of MOGUL is the interdisciplinary nature of the framework, which is designed as a platform to explore (i.e. compare/contrast) modular theories of language from a wide range of disciplines (e.g. linguistics, psychology, cognitive science). Notably, a variety of empirical studies from a variety of disciplines used to inform the framework (Prévost & White, 2000 – feature matching; Marr, 1982 – the visual system; Baars – on consciousness). This creates a research environment where different academic disciplines may be used to inform and evaluate each other. That is to say, research in generative linguistics can inform research in psycholinguistics, which can inform cognitive psychology, and so on. As such, MOGUL attempts to corroborate similar findings from a variety of disciplines into a single harmonized account of the multilingual mind.

Another advantage of MOGUL is that MOGUL is more than just a theory of language or language development; MOGUL is an account of the multilingual mind. While many linguistic theories of language acknowledge that language is part of a larger cognitive system, very few explicate in detail how these systems interact. MOGUL plugs language into a larger architecture and emphasizes the role of extra-linguistic cognitive structures in language production, comprehension and development. The interaction between linguistic and extra-linguistic cognitive systems is particularly relevant to the study of lexical selection (and by extension codeswitching) which may be motivated by any number of non-linguistic factors such as social circumstances or personal goals. For example, this interaction between linguistic and extra-linguistic systems can be seen in the case of lexical selection. When a speaker is choosing a label to describe a militant organization, they will have a number of grammatically sound choices including terms like *freedom fighters*, *rebels* and *terrorists* (VanDijk, 1997). In this case, the label that a speaker chooses may be heavily influenced by a great number of extra-linguistic factors including their personal experience, emotional reactions and even global politics.

Ultimately, the MOGUL platform allows for extra-linguistic aspects of codeswitching (e.g. personal experience) to be accounted for via extra-linguistic cognitive processes (i.e. executive control processes).

Additionally, MOGUL may lead to changes within theories of generative linguistics. Generative theories of syntax seek to explain various linguistic phenomenon by exploiting the grammar. However, as we will see with MacSwan's PF Disjunction Theorem, this may lead to theoretically sound theorems or principles (e.g. a prohibition on intra-word codeswitching) which excludes at least some instances of empirical data—in this case, examples of intra-word codeswitching seen in example (2) above. However, when generative theories are plugged into MOGUL, new mechanisms and operations become available (e.g. cognitive context, goal representations); these new tools made available to generative theories may yield valuable insight and reveal solutions not previously available.

Finally, it is worth noting that MOGUL is a relatively new framework. As such there are still a great number of areas to be explored. Given the grand ambition behind MOGUL (i.e. a model of the multilingual mind), SS&T have appealed to specialized researchers from a variety of academic disciplines to assist in their efforts. In regards to generative linguistics, SS&T state, “[MOGUL] assumes some version of generative grammar but is not committed in detail to any particular one” (SS&T, p. 46). While not all researchers will view this lack of commitment to any one theory as a strength, it provides an excellent platform to evaluate competing generative theories. Answering this call to arms, a crucial goal of this thesis is to provide additional detail to a version of generative grammar which is consistent with the MOGUL framework. To achieve this, this thesis will be evaluating elements of Minimalism and Distributed Morphology within a MOGUL framework in order to account for intra-word codeswitching.

1.1.2. What is MOGUL?

MOGUL is a cognitive framework which is concerned with the real-time representation and processing of language in the multi-lingual mind/brain. Much of the MOGUL architecture is based on work by Jackendoff (1997) on the architecture of the language faculty. Following this perspective, the language faculty is seen as a series of inter-connected modules, each of which performs a specialized function (e.g. the

syntactic module processes syntactic information, the phonological module processes phonological information, etc.)—the details of this framework will be elaborated on in chapter 2.

To develop a clear picture of MOGUL, it is prudent to take a moment to highlight a key aspect which differentiates the two models (i.e. MOGUL vs. Jackendoff’s Model). While Jackendoff focused on the architecture of the language faculty (and what the architecture entails), SS&T are more concerned with the operations involved in online language processing within a Jackendovian style architecture—they are primarily interested in how linguistic representations and cognitive processes construct language in real-time.

Notably, as pointed out by SS&T, although Jackendoff does note that it is possible to interpret his architecture within a processing perspective, he may not endorse the modifications made by MOGUL. For example, MOGUL attempts to re-interpret Jackendoff’s architecture within a real-time (i.e. online) processing perspective as opposed to the abstract ‘time-free’ linguistic perspective (SS&T, 2014) preferred by Jackendoff. Additionally, MOGUL seeks to offer a developmental theory of language based on a Jackendovian architecture and a real-time processing perspective. As such, MOGUL offers a performance-based account of language use which focuses on production and usage of linguistic representations.

1.1.3. Advantages of an Interdisciplinary Framework

One of the major goals of the MOGUL framework is to synthesize a cross-disciplinary and comprehensive view of language processing and development. This cross-fertilization of ideas between disciplines has yielded useful insights into real-time language processing. However, like all interdisciplinary accounts, distinct methodological approaches and ambiguous/conflicting terminology amongst the various disciplines provide challenges that must be reconciled by researchers working in these fields of study. Nonetheless, the rewards for doing so are great; developing a clearly defined working terminology for the MOGUL framework increases the ease with which researchers can cross-examine ideas.

For example, the term *language module* appears in the work of a number of authors from different disciplines. Both Chomsky (1995) and Fodor (1983) define the

‘language module’ as a single organ, analogous to the heart or hand, which is responsible for linguistic processing. However, Jackendoff (1997), in a view adopted by SS&T, envisions a ‘molecular language module’ in which the ‘language module’ consists of a series of modules; some researchers have noted that this view of the language module is reminiscent of a biological system (e.g. the circulatory system). As such, Jackendoff’s model of the ‘language module’ is an amalgamation of several sub-modules which performs the same function as Chomsky’s (1995) module but with a different architecture. These two different ways of understanding the term *language module* are easy to conflate or overlook altogether. While this overlap of jargon is as unfortunate as it is unavoidable, the reward for coalescing interdisciplinary terminology is a stable platform on which we can evaluate divergent approaches to language and cognition. When this thesis encounters such jargon, every effort will be made to provide clear and unambiguous definitions.

While reviewing the MOGUL architecture, it is easy to conflate the MOGUL framework with a hybrid or ‘mix and match’ approach to theory building. SS&T caution against this (Truscott & Sharwood-Smith, 2004; Truscott, 2004), noting that the MOGUL framework is applicable to a number of competing linguistic theories. For example, SS&T make use of Minimalist theory to illustrate syntactic relations in MOGUL but note that other options—for example a Construction Grammar—could also fit within their framework (Truscott, 2004). While daunting at first glance, this inherent flexibility within the MOGUL framework is actually a boon for language mixing researchers as it allows for competing linguistic theories (e.g. Lexicalism & DM) to be compared within a cognitive framework.

1.2.0. Theoretical Challenges for a Cognitive Framework of Language

As pointed out by Carroll (2003), any complete modular explanation of language in the mind/brain will have to satisfy three requirements; a theory of property, a theory of transfer and a theory of learning.

A property theory is required to explain the presence of primitive features in each module. For example, SS&T claim (2014), following Jackendoff (1997), that phonetic representations are symbolic feature representations of real-world objects/events/states (i.e. indirect realism). However, some phoneticians, like Peperkamp (2003, 2004), argue

that phonetic representations are made up of acoustic features which directly reflect their physical parameters in the real world (i.e. direct-realism). As far as MOGUL is concerned, either choice is acceptable; what is important here is not the choice of content but rather that modular properties are explicitly stated and the recognition that the properties contained in each module are fundamentally different.

In any modular account of the mind/brain, a theory of transfer is needed to explain how representations in each module interact with each other. Emanating from Fodor's notion of 'vertical modular organization' (1983), the central concern here is how isolated modules can interact and pass information throughout the computational system. This can be formally restated as: how much of level X-1 can level X see. For example, in the Minimalist Program introduced by Chomsky (1995), syntactic structures (i.e. the syntax module) directly interface with semantic structures. Additionally, the syntax module also interfaces to phonological representations during PF spell-out. However, there is no direct interface between semantic and phonological structures. Thus, in this account, the syntax module can peer into the conceptual module and the phonological module, but the phonological module can be isolated from the conceptual module. Once again, what is important here, for any modular theory of language, is not what the specific modular relations are but rather that a set of relations is required.

Additionally, a theory of learning is needed to explain how new linguistic forms are created/acquired. Language and linguistic representations are not static; they are in a constant state of flux. Children acquire grammatical forms and lexical items regularly and adult second language learners are able to develop new phonetic categories and syllable templates. For example, a Mandarin speaker studying English will have to learn how to produce complex onset consonant cluster (e.g. CCCVC as in 'strong') which do not appear in their L1. As such, any framework attempting to account for how language resides in the mind/brain will have to be able to explain how language and linguistic representations can change.

The MOGUL framework is able to satisfy Carroll's three requirements. As a theory of property, MOGUL adopts the linguistic notion of a Universal Grammar and generative linguistics as presented by Chomsky (1995) and championed by Fodor (1983). UG postulates a set of universally innate primitive features which reside in a language

module. These primitives are the building blocks of natural language and are exploited by language learners. UG has been traditionally employed by generative linguistics to account for a puzzle known as the *Poverty of the Stimulus* (Plato's Problem)—the notion that in natural language learning, learners receive insufficient data to account for learners converging on a complex set of rules and that innate features of language must be postulated to account for this fact. By exploiting UG as a theory of property, MOGUL is able to epistemologically account for the innate primitive features found in each module. The exact shape of these features and their role in building representations is the subject of some debate and will be discussed in greater detail in section 2.1.0.

The MOGUL framework views Acquisition by Processing Theory (APT – SS&T 2004) as both a theory of transfer and a theory of learning. APT is a central tenet of MOGUL and will be reviewed in detail in section 2.3.0. For now, let it suffice to note that in MOGUL, information is not ‘transferred’ in the metaphorical sense but rather modular specific representations are chained (i.e. co-indexed) to representations in other modules. There is no formal Language Acquisition Device (i.e. LAD); instead SS&T claim that as these chains are developed through experience and strengthened via frequency of use—this strengthening of representations equates to learning.

1.3.0. Chapter Summary

In summary, this thesis will use a MOGUL processing perspective to build upon previous accounts of language mixing to account for intra-word codeswitching in bilingual speech. The interdisciplinary nature of the MOGUL framework makes it the ideal platform to develop a detailed account of intra-word codeswitching; additionally, the phenomenon of language mixing will be used to support the adoption of a specific theory of generative grammar into the MOGUL framework. Notably, this thesis will accept much of MacSwan's Minimalist account while rejecting the PF Disjunction Theorem and appealing to elements of Distributed Morphology (i.e. late insertion and lexical decomposition in the syntax). This thesis will argue that cognitive context interacts with elements from executive control (i.e. goal representations—central motivators for thought and action like *satisfy hunger* or *be funny*) which will in allow for the union of phonological systems between Language X (L_x) and Language Y (L_y). This analysis builds upon a larger body of language mixing research by synthesizing a DM

account of codeswitching with a cognitive processing framework to account for intra-word codeswitching. Ultimately, it will be argued that intra-word codeswitching is a natural product of the MOGUL framework.

Chapter 2: The Cognitive Framework

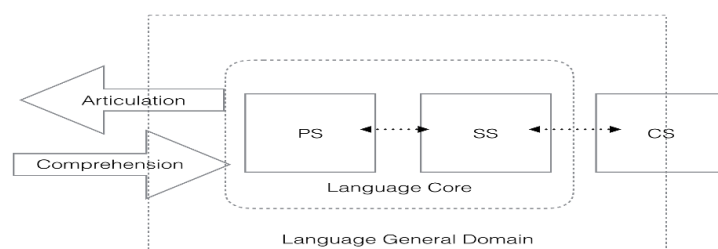
2.0.0. Cognitive Framework

This section will review the necessary components of a cognitive framework (i.e. cognitive components or mechanisms) which will be exploited to account for intra-word codeswitching. It is worth noting that this is not a comprehensive or exhaustive description of MOGUL (SS&T, 2014). While MOGUL works to provide a complete and coherent sketch of language development in a multilingual mind, due to time and space considerations I will focus on cognitive operations that are intimately involved in codeswitching.

The first half of this chapter will begin with a general review of the MOGUL architecture as well as a detailed discussion of the MOGUL fundamentals (i.e. representation & processing). The latter half of the chapter will discuss key cognitive operations (i.e. cognitive context & executive control processes) and how they fit within MOGUL.

2.1.0. MOGUL Architecture

Developed by Sharwood-Smith and Truscott (2014), the MOGUL framework offers a modular model of the organization of language in the mind/brain. The architecture of MOGUL draws heavily on Jackendoff's tripartite model of language (Jackendoff, 1997, 2003) where linguistic faculties (i.e. phonology, syntax) are encapsulated modules—in the sense of Fodor (1983)—and are independently interfaced with other linguistic and extra-linguistic modules (e.g. conceptual module). This architecture is pictured below in fig. 1.

Figure 1: Tripartite Model

(Adapted from SS&T, 2014)

Crucially, each module (also referred to as *structures* as in syntactic structures (SS) and conceptual structures (CS)), contains its own unique set of primitive features which are assembled to form representations from linguistic input. For example, the SS (Syntactic Structures) may form an SS representation like [+noun, +nominative, +singular, etc.] for the word *cat* in the phrase *the cat is evil*. Likewise, the PS (Phonological Structures) will independently construct a phonemic representation from phonological primitives—perhaps something akin to Distinctive Features (Jakobson, Fant, & Halle, 1951). Once each module constructs its own representation out of the set of primitives available to it, the representation is then interfaced to its neighboring modules. The result of this interfacing is a (PS + SS + CS) representational chain which contains all the necessary phonemic, syntactic and conceptual information to be articulated as a word. This framework and its component processes will be explicated in greater detail below. I will argue that language mixing is the result of constructing representational chains using features from Language-X (L_x) and Language-Y (L_y).

2.1.1. Basic Architecture

For purposes of exposition, we will divide modules into two groups: linguistic modules and extra-linguistic modules. Linguistic modules include the phonology module and the syntax module. Crucially, these two modules are specific to language processing and constitute what we will call the *language core* (see fig. 1). As noted above, the language core is encapsulated in the Fodorian sense (1983) and is restricted to interfacing with a limited number of extra-linguistic modules. The nature of these inter-modular

relations will be discussed below. Extra-linguistic modules include the perceptual modules, the motor-control module and the conceptual module. While these modules are involved in language processing (i.e. semantics, speech perception and production) they are *extra-linguistic* in the sense that they are part of a general cognitive apparatus that governs action and knowledge beyond just language. These modules will be collectively referred to as *language-general domain*. The conceptual module, which plays a central role in establishing word meaning, is of particular interest and will be discussed in greater detail below. The terms *structure* and *module* will now be used interchangeably (note that the phonological module will, following SS & T, be abbreviated as PS, and the syntactic module as SS, etc.).

The modular model of processing presented by Sharwood-Smith and Truscott draws heavily on the previous work by Jackendoff (1999). The language-general domain (e.g. lexical conceptualization and perception) surrounds the language core and contains modules used in processing both linguistic and extra-linguistic information (i.e. any information which is additional to the linguistic message). The language core (e.g. phonology and syntax) can interface with modules in the language-general domain (conceptual module) which in turn are interfaced to a greater network of modules (e.g. olfactory module, kinesthetic module, etc.).

2.1.2. Inter-modular Relations: General

The model SS&T propose consists of several independent modules, each with its own unique integrated processor and information store which are, in turn connected to other modules by interface processors. A detailed explanation of these components will follow in section 2.2.0.

The language-core is made up of two language-specific modules: phonological structures and syntax structures. These encapsulated modules do not have direct access to external stimuli but rather take the outputs of extra-linguistic structures as inputs (e.g. CS). The language-core is interfaced with more language-general modules (i.e. extra-linguistic modules) which are part of general processing (i.e. the conceptual module or the perceptual structures). Generally speaking, these interfaces between the language-core and the language-general domain are responsible for sound-meaning

correspondences (i.e. morphemes, words, phrases)³. This process will be explicated in detail below in the subsequent sections.

2.1.3. Inter-modular Relations: Key Modules

Of central interest to MOGUL are the Phonological Structures (PS) and Syntactic Structures (SS) which constitute the language-core domain, in addition to the Conceptual Structure (CS)—the module responsible for conceptual representation. Fig. 1, shown above, illustrates this tripartite set of modules in MOGUL. Following Jackendoff (1997, 2003) SS&T claim that the PS and SS directly interface with each other and constitute the familiar linguistic notion of a language module (despite being internally complex themselves).

2.1.4. Inter-modular relations: PS < - > SS (The Language Core)

The interaction between PS and SS can be seen when considering word stress in an example like *record*. If the PS constructs the representation [ˈrɛk ərd], the PS-SS interface will link the PS representation with the [+noun] SS representation, since in English, the word stress associated with the PS string in question is indicative of the noun. However, if the PS forms the representation [rɪ ˈkɔrd], the PS-SS interface will link the PS representation with the [+verb] feature in the SS representation. The nature of features contained in each information store will be discussed in greater detail in section 2.2.1.

Crucially, SS&T (2016) propose that in MOGUL the language core operates blindly—without knowledge of which language/dialect/register is being processed. This is a natural consequence of the MOGUL architecture. While this ‘meta-knowledge’ of language is available in the CS (see section 2.4.5), MOGUL views the language core as ‘neutral territory’. “[The] PS and SS operate efficiently and blindly with any relevant input that appears in their interfaces” (SS&T, 2016: p. 6). In other words, SS and PS structures are not tied to any one language but rather formed from an assortment of universal primitives which are accessible by all languages. However, through linguistic experience, which associates linguistic representations with the contexts which activate

³ Generative theories, like DM and Minimalism, argue that semantic meaning is part of the language core; in MOGUL ‘meaning’ is extra-linguistic.

them, a language user will come to identify and associate a specific set of linguistic elements to a specific language (or dialect or register). Notably, it is this ‘blindness’ in the language core that allows for elements associated with language A to appear in language B in this account of intra-word codeswitching. This point will become clearer in section 2.2.0 as this account of processing in MOGUL is further developed,

2.1.5. Inter-modular relations: SS < - > CS

Fig. 1 also highlights the direct relationship the language core has with extra-linguistic modules. Of particular interest is the CS which can directly interface with both the PS and SS. While representations in the CS may be generally thought of as something akin to ‘word-meaning⁴’, the CS representations are better understood as abstract features. Jackendoff (2005) refers to CS representations as “quasi-algebraic” representations of lexical items. These representations are highly abstract and are not consciously accessible to an individual—they are not part of *lexical* knowledge. Crucially, this means that CS representations may be shared by multiple words or phrases (e.g. cheap/thrifty). This point will be expanded upon in section 2.2.4 when we discuss how extra-linguistic modules work with the CS to establish a semantic chain of representation.

To exemplify CS representations, SS&T highlight theta-roles (i.e. agent or patient) as an illustration of the type of content in the CS information store. This allows for CS representations to be independently linked to syntactic structures. For example, let us consider the following sentences

- A) The police chased Jack up a tree.
- B) Jack paid off the police.

In examples (A) and (B), the word *police* is represented in the SS as [+noun]. Additionally, in (A), *police* occupies the subject position and will also contain the feature [+Nominative] as part of the SS feature bundle which will be linked to the [+agent] feature in the CS (i.e. [SS(+noun, +nominative) + CS(+agent)]). However, in (2), *police* is assigned the [+accusative] feature in the SS which will correspond to the [+patient] feature in the CS (i.e. [SS(+noun, +accusative) + CS(+patient)]).

⁴ In MOGUL, the CS is the locus of meaning for morphemes, words, phrases and sentences.

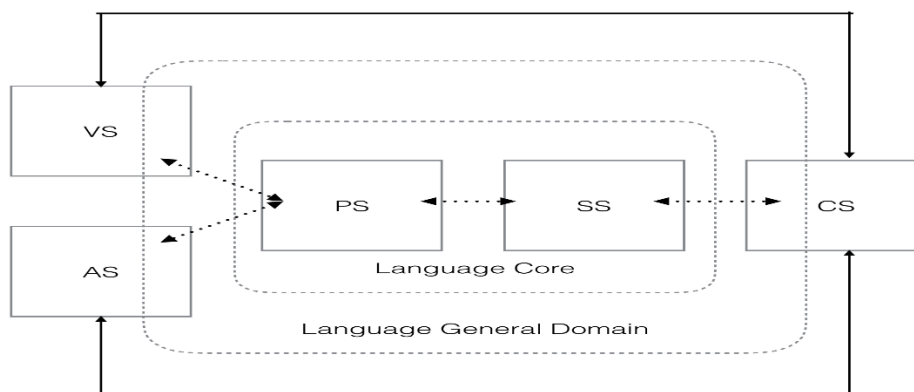
SS&T note that we do not necessarily have conscious access to our CS representations (i.e. [+agent]) when consciously considering the meaning of a concept (e.g. morpheme, word, phrase) but rather we access a semantic chain of inter-modular representations and associations (i.e. POpS + AffS + etc.).

Additionally, the CS contains a ‘general language representation’ (GLR); this GLR is triggered by context and co-indexed with all representations associated with the language it represents. This point will prove crucial when accounting for language selection in the MOGUL framework and will be expanded upon in section 2.4.5.

2.1.6. Inter-modular relations: AS < - > PS

Extra-linguistic modules such as the Audio Structure (AS) and Visual Structure (VS) also play a significant role in representing word-meaning. The AS and the VS denote the aspects of our sensory perception that have an influence on language. SS&T distinguish between an AS representation containing no linguistic information (e.g. the sound of a bell or a car backfiring) and an AS representation containing a linguistic message (e.g. hearing a word); the latter is labeled AS_L. An AS_L representation contains both extra-linguistic information—including features which indicate gender, age, etc.—as well as information encoding the phonetic structure of the word. As this thesis is primarily concerned with language and language use we will use the abbreviation AS to stand for AS_L.

The AS-PS interface allows for direct correspondence between acoustic and phonemic representations; PS representations consist of symbolic structures representing linguistic sounds (i.e. something akin to distinctive features). For example, if a listener hears the word *cat*, the acoustic stimulus enters the audio system: the AS constructs a phonetic representation for the input and creates an output; the AS output interfaces the linguistic stimulus to the PS and becomes the intake for the PS. The PS then sets up a phonemic representation - [kæt] - based on the intake. This relationship allows extra-linguistic information found in the AS (i.e. pitch, amplitude, etc.) to be separated from the linguistic information in the sound signal.

Figure 2: Key Relation between Modules

(Adapted from SS&T 2014, p. 17)

2.1.7. Inter-modular relations: AS < - > CS

Meanwhile, the AS-CS interface links extra-linguistic information found in the stimulus to the CS; this provides the listener with a plethora of extra-linguistic information about the stimulus, including but not limited to the origin of the signal (e.g. man, woman, machine), the emotional state of the speaker (e.g. angry, excited), and even the shape of the environment (e.g. acoustic feedback/echolocation). As such, AS-CS interface accounts for the way alterations in pitch and amplitude relate to conceptual structure⁵, namely, revealing extra-linguistic information about speaker such as physical size, age, gender or even location⁶.

2.1.8. Inter-modular relations: VS < - > PS

Likewise, the VS-PS interface accounts for orthographic relations to sound as well as visual influences on speech perception as demonstrated by the McGurk effect (Kawase, Hannah & Wang, 2014). In a classic example of the McGurk Effect, a participant is played an audio clip of the phoneme [b] while simultaneously watching a video clip of a speaker producing the phoneme [g]. The result in the participant

⁵ Note that, in MOGUL, there is no purely linguistic semantic level of representation; conceptualization of meaning is extra-linguistic.

⁶ Information about size, age, gender, etc., is part of the mental image a listener develops to represent the speaker; in MOGUL this is realized in POpS via the process of synchronization (see section 2.1.9 for details). For example, an AS representation of a low pitch speech signal may be synchronized with other POpS representations to form a mental image of a large male.

perceiving the phoneme [d]—a merger of audio and visual stimulus. The McGurk Effect clearly demonstrates multiple channels of perceptual input. In MOGUL terms the McGurk Effect is seen as the result of competition between the AS-PS and VS-PS interfaces where the PS seeks to create a coherent representation based on conflicting intakes.

2.1.9. Inter-modular relations: CS < - > POpS & AffS

As noted above in section 2.1.5 (SS<>CS), CS structures consist of abstract primitives (e.g. theta roles) or combinations thereof. However, the CS interfaces with a number of other extra-linguistic modules and forms representational chains which account for complex word meaning. Central to developing these complex meanings is sensory perceptual system. This system consists of five independent modules, referred to as Perceptual Output Structures (henceforth – POpS), each accounting for a unique form of sensory input (i.e. taste, touch, sound, smell and sight). In addition to the Audio Structures (AS) and Visual Structures (VS), POpS also contains Gastronomical Structures (GS), Olfactory Structures (OS) and Tactile/Kinesthetic Structures (TS)—these modules are illustrated below in fig. 3. While OS, GS and TS have a minimal role in language processing, they are crucial in establishing complex semantic and pragmatic meanings by interfacing with the CS.

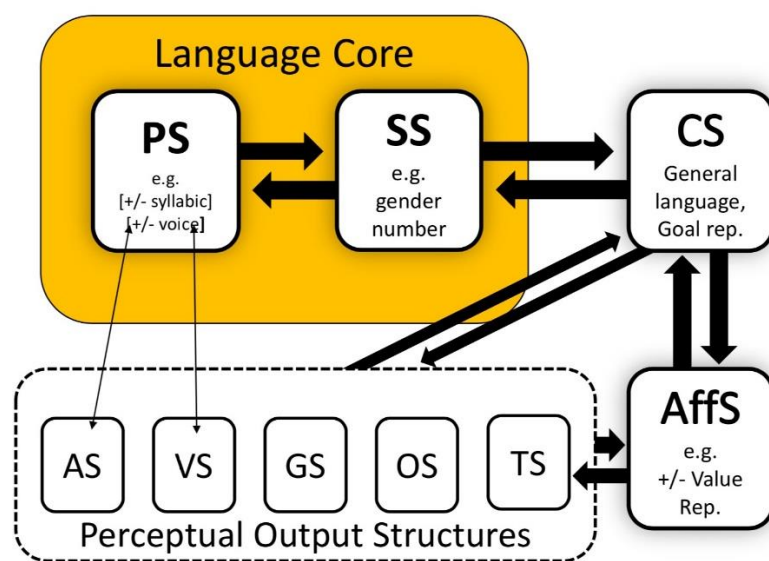
These sensory perception modules undergo a process called POpS synchronization, by which structures within POpS are synchronized to form a uniform representation. This uniform representation contains all sensory information about the element being represented and constitutes what we may think of as a ‘mental picture’. As pointed out by SS&T, it is important to note that these perceptual representations do not directly correlate to a physical stimulus but rather represent the experience of said stimuli. As such, a representation within the VS should not be thought of as “internalized image in any simple sense” (p. 146), but rather something more akin to a digitized image (e.g. a .jpeg image); likewise, other representations in POpS will form representations in a similar fashion using their own set of primitives.

For example, if an individual is asked to think of a cow; their visual structure will provide a visual representation of what the animal looks like; their AS will sync with the VS and provide the visual representation with a co-indexed AS (e.g. cow sounds);

additionally, representations from the OS (e.g. manure smells) and TS (e.g. coarse hair) will combine with the VS and AS to complete the ‘mental picture’ of a cow before interfacing with the CS.

It is worth noting that no GS feature is synced to the mental image of ‘cow’. This is because, at least in English, people do not eat ‘cow’; people eat ‘beef’, leading to a conceptual distinction. Nonetheless, these two concepts are intimately related—as beef is cow flesh the concepts are closely associated with each other which means activating one will tangentially activate the other via the process of spreading activation (see Rumelhart & McClelland, 1981). While the GS structure for beef (i.e. savory red meat flavor) is not synchronized with POpS representations for ‘cow’, this associated representation is activated and will play a vital role in developing cognitive context—this will be touched on below in section 2.2.4 and fully explicated in 2.4.5, where the notion of cognitive context will be further developed and shown to play a central role in language selection.

Figure 3: MOGUL Architecture



(Adapted from SS&T, 2014: p. 161)

Additionally, MOGUL accounts for emotional influences via an affective module—labeled AffS. Similar to POpS, the AffS is a part of general cognitive processing and not specific to language. These AffS structures are responsible for accounting for both simple and complex emotions that operate both above and below conscious awareness. They do this by assigning a value feature to interfaced (i.e. co-indexed) representations; high value representations correspond to elevated levels of

activation for representations that are part of the same chain. As SS&T explain, “These AffS structures have the effect of assigning a more or less positive value or a more or less negative value to the representations they are co-indexed with and may be set or reset at any given moment” (2016: p.3). These AffS structures are co-indexed to POPS and complex CS representations and constitute the affective or emotional aspects of meaning⁷.

For example, let us consider the words *cheap* and *thrifty* in the context of *Sam is ____*. Conceptually, these terms are viewed as nearly synonymous, differing only in that *thrifty* has a positive value representation in the AffS, while *cheap* has a negative one. Thus, the semantic chain of representations (i.e. CS+POPS+AffS) for these two lexical entries may be nearly identical, differing in that “thrifty” would be represented as [+positive] in the AffS while *cheap* would have the [+negative] AffS representation. As such the selection of one term over the other may be emotionally motivated; the choice of whether Sam is ‘cheap’ or ‘thrifty’ may depend on whether or not the speaker likes Sam. Crucially, SS&T have noted the central role AffS structures play in both lexical selection and codeswitching; this will be expanded upon below in section 2.2.4 (lexical selection) and 4.2.0. (MOGUL & codeswitching)

The key take-away from this discussion is the formation of complex meanings in the CS. As noted above, CS primitives are abstract and below the level of conscious awareness; the co-indexation of POPS and AffS structures to CS structures creates a complex representational chain (i.e. CS + POPS + AffS) which constitutes the semantic and pragmatic aspects of meaning in a word. As such, for the remainder of this thesis, when ‘CS’ representations are being discussed they should be thought of as complex CS chains which incorporate POPS & AffS structures.

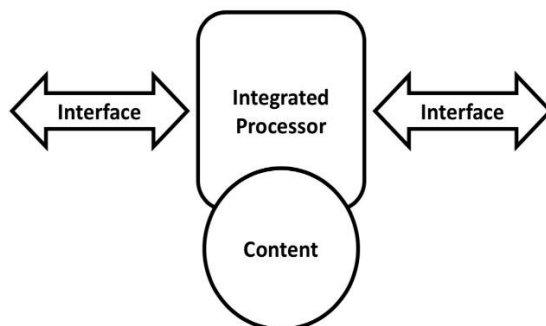
2.2.0. MOGUL and Modularity

Modularity is a key component of the MOGUL framework. As stated above, in section 2.1.0, each module in the MOGUL architecture consists of an integrated

⁷ In MOGUL, emotion and affective structures are much more complex than presented here; the account of affective structures presented in this thesis is meant to highlight value representations and their role in language selection. How complex emotions (e.g. jealousy, avarice, unease) are represented in MOGUL is beyond the scope of this thesis.

processor as well as an information store. The modules are linked to each other via interface processors. The basic components of each module are illustrated below in fig. 4.

Figure 4: Diagram of a Module in MOGUL



Crucially, all modules, regardless of their function, have the same internal architecture—the PS (phonology structure), AffS (affective structure), and VS (visual structure) all consist of an integrated processor and an information store which are connected to other modules via interface processors. While each module will contain its own set of primitive features and be interfaced to a unique series of modules (i.e. inter-modular relations), the fundamental structure of each module is the same.

2.2.1. Information Stores & Representations

The content found in information stores is made up of coded primitive features which are accessible for processing; these primitives (or combinations thereof) are unique to each module and can only be manipulated by a module-specific (integrated) processor to form a module-specific representation for a given input. This means that primitive features which are present in one module will not exist in another. For example, SS&T claim that primitive features in the syntax module's information store may include items like [+noun] and [+tense] which do not exist in any other information store, while the phonological store might contain distinctive features such as [+strident], [+continuant] or [+voiced].

Content in the information stores is not static; it encodes what SS&T call an *activation level*. The activation level of any feature in an information store is constantly in flux. When static, content in the information store will sit at a resting level of activation. As all elements in the system have mental associations, and associative connections are

dynamic (i.e. they co-activate each other) all elements will have a resting level of activation based on previous usage as well as the strength of its associations. As stated by SS&T, “an item’s activation level is commonly seen as a function of its past use in processing and therefore reflecting learning by the system” (p. 68). The notion of ‘learning by the system’ and its effect on resting activation levels will be discussed in greater detail below in section 2.3.0 (APT).

Additionally, the activation level of an element in the information store is affected by ‘spreading activation’ between associated elements (Rumelhart & McClelland, 1981)⁸. For example, SS&T (2014) note that a listener’s interpretation of the term *bank* will be affected by whether or not they just heard the term *river* or *money*. This occurs as the terms *river* and *bank* each prime a specific meaning of the term *bank*—that is, either, the side of a ditch along a river or a secure location to deposit one’s money. However, as the two senses of the term *bank* are homonyms, they also phonologically prime each other (SS&T, 2014). So, the term ‘bank’ will in isolation cause the processor to represent both senses of the word, but collocation with another term, such as *money*, will increase the activation level of a particular sense of *bank*—in this case, let us say a financial institution—which will cause it to win the competition.

2.2.2. Integrated Processors & Active Content

Within each module, the integrated processor (also referred to simply as processors) is responsible for constructing a representation for any active content in its information store. Content becomes active when triggered by a stimulus, be it external (reading or listening) or internal (from the CS). Stimuli that feed integrated processors will be referred to as *input*; the representations produced by integrated processors feeds interface processors and will be henceforth referred to as *intake* (see Carroll, 2004, for details on input processing). As input/intake are parsed, a number of corresponding primitive features will become active in each information store. As the processor activates combinations of features in the information store, multiple representational

⁸ The notion of ‘spreading activation’ was the source of much debate in linguistics (see Pinker & Ullman, 2002; McClelland & Patterson, 2002); however, ‘spreading activation’ is supported by in Cognitive Psychology (e.g. Foster, Hubbard, Campbell, Poole, Pridmore, Bell, & Harrison (2017) - on the spreading of activation in non-verbal/visuospatial memory networks)

constructs (i.e. bundles of features) will compete to become the successful representation. Ultimately, the representation is formed from the bundle of features which has the highest level of activation at the time of processing; this representation is the output of the integrative processor and will subsequently be the intake for the adjacent interface processors.

For example, let us consider how the PS constructs a representation for the word *record* if an individual were to read the phrase "...*the new record*...". The input would enter the VS and incrementally receive (i.e. from left to right) an orthographic representation as part of the feature bundle representing each element in the phrase. When the parser reaches the term *record* the VS-PS interface will take the VS representation (i.e. the symbols r,e,c,o,r,d) as its intake. The PS processor will then construct two competing representations: [ˈrɛk ərd] referring to a noun and [rɪ ˈkɔrd] representing a verb. If the word *record* is read in isolation, the PS would have no way to determine which representation best matches the intake the choice between the two competitors would be arbitrary. However, in our example the determiner *the* introduces the term *record* which is indicative of a noun. As such, the presence of a determiner preceding the term *record* will increase the activation levels of features in the nominal representation, [ˈrɛk ərd], which, in turn, will cause [ˈrɛk ərd] to be selected as the PS representation over [rɪ ˈkɔrd].

In many situations, the representations produced by the processors are not novel; if an input/intake has been experienced before, its pre-existing representations and chains will be retrieved from long-term memory (see section 2.2.5 for details). However, if the input/intake are novel the processor may construct a novel representation; if no current representation for an input/intake is found in the information store a new one will be established by modifying or combining existing structures. It does this by creating several preliminary representations which will compete within a module in the manner described above. The representation in the integrated processor which becomes most active will then function as the intake of the interface which co-indexes the most active representations to other modules.

As a final point, it should be noted there is no activation threshold (SS&T, 2014: p. 75) that content must cross to become a representation; rather it is merely the most

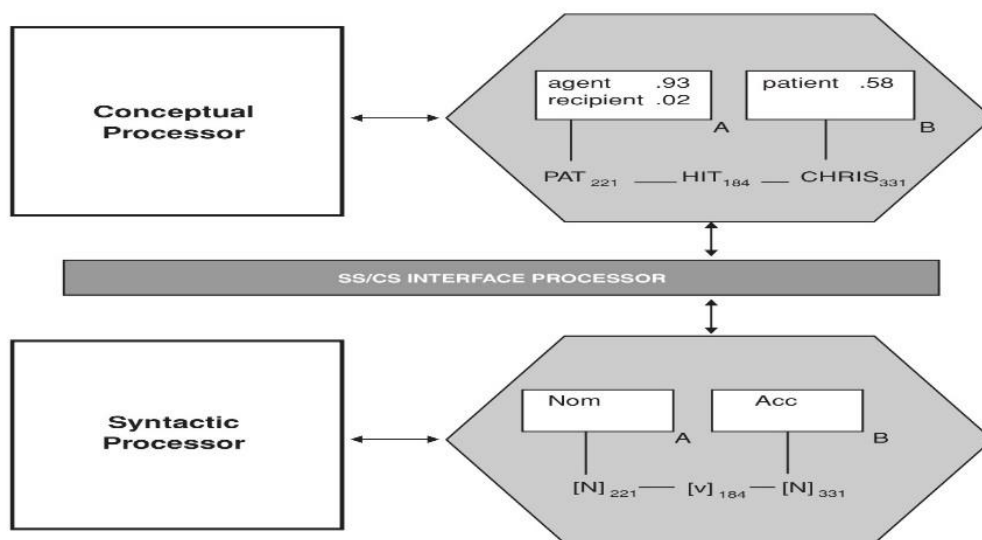
active content that becomes the intake for the interface processor. As such, the activation levels of any content in an information store is constantly in flux as interface processors attempt to match intakes from various processors. Ultimately, it means that intake representations are continually updated or modified throughout a processing event as new elements are parsed.

2.2.3. Interface Processors, Co-indexation & Chains

The output representation produced by each integrated processor constitute the intake for the interface processors (hereafter referred to as *interfaces*). The processor takes any active primitives (i.e. content) in the information store and arranges/combines them to create a representation of the input/intake which is then linked (or chained) to representations of the same input/intake generated in other modules by their processors. These interfaces match the most activate intakes from neighboring modules in a process referred to as *co-indexation*. As noted above, in section 2.1.4 [PS \diamond SS], these interfaces are blind to the content they are co-indexing and merely match the most active content in each module. Intakes from two or more modules which are co-indexed from a chain of representations. Like integrated processors, interfaces construct chains out of the representations available to them, which also compete for selection.

If the interface has encountered the intake items before it may use a previously established chain stored in long-term memory; if the intake elements are novel then the interface will construct a novel chain or modify a pre-existing one. SS&T state, “[An interface processor’s] function is to match activation levels of adjacent modules and assign indexes to new or existing items as a necessary preliminary to activation matching” (SS&T, 2014: p. 39). The result is a chain of representations (i.e. PS+SS+CS) which contains phonological, syntactic and conceptual information; this mental chain of representations constitutes the traditional linguistic notion of ‘word meaning’.

Figure 5: Diagram for ‘Pat hit Chris’



(SS&T, 2004: p. 10)

For example, consider a SS-CS chain for the English sentence *Pat hit Chris* as presented by SS&T (2004) in fig. 5 (where the hexagon = information store) which is meant to highlight competition between theta-roles. When a language user encounters this sentence the SS will begin to construct a representation for the first element—in this case *Pat*. Based on previous experience (i.e. established grammatical principles) the SS will first assign the nominative case to the first NP. This will increase the activation level for [+nominative] in the SS which will cause the SS-CS interface to co-index the SS feature with the highly-correlated CS feature, [+agent], increasing its activation level. Note, however that at this early stage in the computation there is more than one option for theta role assignment to *Pat*; for example, in the sentence *Pat received a gift*, *Pat* is the recipient of the action. These activation levels are reflected in fig. 5 by the numeric value assigned to each potential theta role (agent - .93 / recipient - .02).

As such the activation level of [+recipient] will compete for selection with [+agent] in the CS. Following our example in fig. 5, the feature [+agent] has a greater degree of activation and will win the competition. As the computation proceeds, representations will be incrementally constructed for the elements *hit* and *Chris* in a similar fashion and the sentence *Pat hit Chris* will be the ultimate output. However, if the noun *Pat* is followed by a passive verb phrase like *was hit*, the activation levels for

Pat [+agent] will decrease while the [+ patient] feature will receive a boost in activation. This incremental processing not only re-adjusts activation levels for *Pat* but will also affect the features assigned to *Chris*; as the activation of [+ patient] will increase for *Pat* while the activation of [+agent] will increase for *Chris*.

2.2.4. Conceptual Representations, Metalinguistic Knowledge & Lexical Meaning

As already noted, conceptual representations and extra-linguistic knowledge are central elements in a MOGUL account of codeswitching; these elements work together to form lexical meanings in the form of complex CS representations. The following discussion will review key points we have discussed so far and provide an example of the formation of lexical meaning.

As mentioned above in section 2.1.5, the CS exists outside the language-core and is heavily influenced by extra-linguistic modules during the processing of any given intake. Thus far, we have introduced Audio (AS) and Visual (VS) structures. Within a MOGUL architecture, these two modules are part of a larger amalgamation referred to as Perceptual Output Structures (POpS) which contain all five sensory-perception modules (i.e. also olfactory, gastronomical and tactile). As pointed out by SS&T, it is important to note that these perceptual representations do not directly correlate to a physical stimulus but rather represent the experience of said stimuli. SS&T also point out that there is a high degree of what they call *synchronization* that goes on in POpS; representations within each sensory-module are co-indexed in POpS and merge to form a synchronized representation. In other words, a visual stimulus represented in the VS synchronizes with representations in AS, GS, OS and TS/KS which chain together to account for the descriptive aspects of lexical knowledge to which we have conscious access.

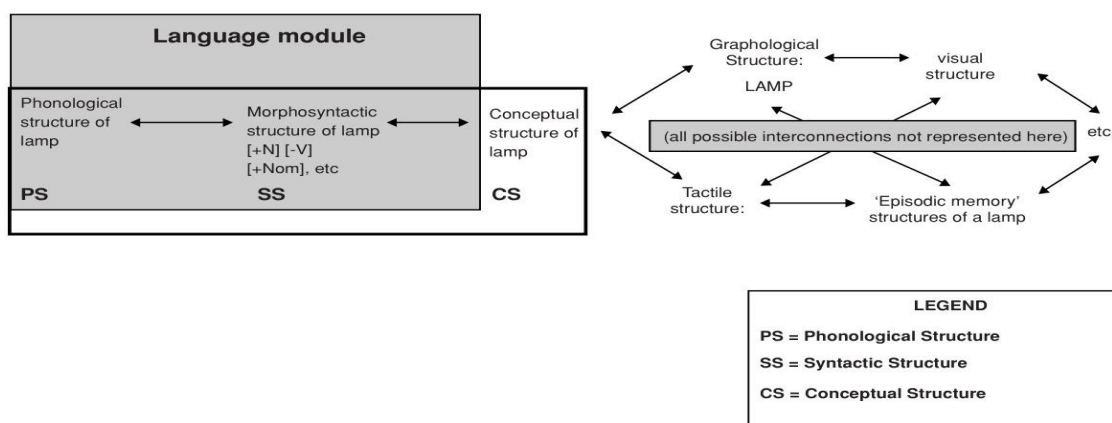
For example, let us consider the lexical entry *lamp*. Within the language module, the PS will construct the phonological representation [læmp], while the SS will combine active features into feature bundles to construct its own representation. External to the language module, the CS constructs its own representation for *lamp* in the manner detailed above (section 2.1.9). However, this alone is insufficient to account for complex perceptual and affective associations linked to the concept ‘lamp’. The CS must interface its representation with additional intakes from perceptual modules to complete the

semantics of *lamp*. Notably, as perceptual and affective associations are unique to each person (developed via personal experience), representational chains for the concept of *lamp* will vary from individual to individual.

As illustrated below in fig. 6, the perceptual and associative information emanating from POpS (i.e. VS & AS) and other extra-linguistic operations (e.g. context, episodic memory) merges to form the conceptual representation of *lamp*. Notably, the representations that constitute the lexical meaning of *lamp* are not CS representations but rather are representations formed in extra-linguistic modules that are co-indexed to the CS.

While this thesis has introduced and discussed several extra-linguistic modules which are particularly relevant to this story, the CS is also co-indexed to numerous additional modules which are beyond the scope of this thesis but still contribute to lexical meaning (e.g. episodic memory, operational knowledge, logic, etc.). These relations are demonstrated in fig. 6 where the [PS + SS + CS] chain for ‘lamp’ contains a complex CS representation which is co-index to several extra-linguistic representations.

Figure 6: Conceptual Representation for Lamp



(SS&T, 2004: p. 3)

Additionally, one other module is required in MOGUL to account for the role of emotion in linguistic processing: the Affective Structure (AffS)—see section 2.1.9 for details. AffS structures are co-indexed to the CS in the form of value representation

which accounts for the emotional influence of CS representations co-indexed to the language core⁹.

Together, the POpS and AffS systems play a major role in establishing lexical knowledge. Our conscious understanding of a word's meaning is the amalgamation of our perceptual understanding (POpS), affective reaction (AffS) and conceptual structure (CS). In other words, MOGUL accounts for lexical meaning via a (CS + POpS + AffS) chain of representations. For ease of exposition, the remainder of the thesis will consider lexical meaning to be part of a complex CS representation unless otherwise stated.

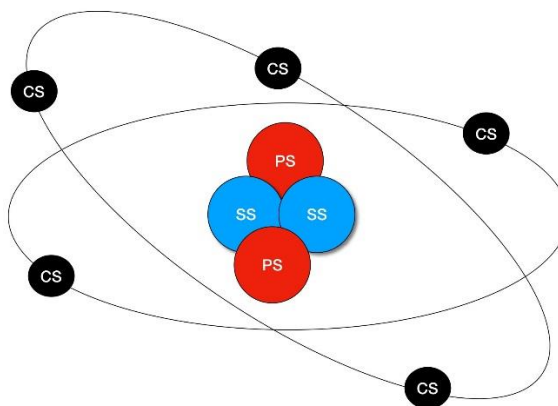
2.2.4. Lexical Entries, the Lexicon & Memory

By now, a reader may have already realized that the picture of the Lexicon being sketched out by MOGUL looks quite different from traditional (albeit varied) approaches in generative linguistics (i.e. the lexicalist hypothesis – MacSwan, 2005). Let us take a moment to fill out the MOGUL view of a lexical entry and the lexicon.

The lexicon consists of an array of lexical entries. In MOGUL, a lexical entry is not a single representation but rather an amalgamation of module specific representations that are both external (CS + POpS + AffS) and internal (PS + SS) to the language module. Specifically, a lexical entry is a chain of co-indexed representations [PS + SS + CS(POpS + AffS)] (see fig. 6 above). Metaphorically, this representational chain presents a molecular view of a word, pictured below in fig. 7. Just as the nucleus of an atom (i.e. the language core) is made up of protons (i.e. SS) and neutrons (i.e. PS) and surrounded by a number of electrons, this metaphor is meant to demonstrate how disparate features from each module come together to form a word.

⁹ AffS are episode dependent (i.e. variable)

Figure 7: The Molecular Metaphor: A word as an atom/molecule¹⁰



Under the MOGUL view, dormant (i.e. non-active) lexical entries are stored in long-term memory. When a lexical chain in long-term memory becomes active it is written in ‘working memory’ (SS&T citing Jackendoff 1987, 1997) where the chain is subject to lexical competition and manipulation by interface processors. These two levels of memory reflect two states of activation—a resting level of activation and an active/current level of activation—that will be discussed in greater detail below (section 2.3.0 APT). Crucially, SS&T note that working memory is not a place (i.e. cognitive module); an item in working memory is merely an active item in long-term memory (2014). They state “[working memory is] a transient pattern of activation of elements within long-term memory stores, possibly in novel configurations” (SS&T, 2004: p. 3). As such, the lexicon in MOGUL is merely a sub-set of highly structured long-term memories which contain patterns of activation (i.e. co-indexations) for feature bundles in multiple modules. In other words, the lexicon consists of chains of representations (i.e. lexical entries) stored in long-term memory.

2.2.5. Lexical Access: Dual Storage & The Processing Race

A model of language and the mind/brain requires a theory of lexical access—a method for accessing stored lexical elements during language processing. A central concern when discussing lexical access is whether lexical elements like *sadness*, *refrigerator* and *dogs* are stored as whole chunks or broken down into their

¹⁰ It has been pointed out to me that the term *atomic* is often used to represent the smallest possible unit available (despite sub-atomic structure); as such, to avoid confusion, the term *molecular* is used to highlight the interconnected nature of separate elements even though fig. 7 is a clearly an atom.

compositional elements (i.e. morphological elements). While it is easy to see the computational nature of some elements like *trees* (i.e. tree[N] + s[+plural]) it seems much less plausible that a word like *refrigerator* is mentally stored as ‘refrigerat(e)’[v] + ‘or’ [+agent], given the low frequency for the verb *refrigerat(e)*. Thus, it may seem intuitive to claim that *refrigerator* is stored as a whole chunk; however, this treatment is unsatisfactory for terms like *trees* which would be required to have separate lexical entries for *tree* and *trees*.

There is a third possible way to account for lexical access by combining the first two. In the view espoused by MOGUL, both whole-word chunks and the component parts are stored in memory, and lexical items may be accessed by activating either the unanalyzed whole form or by the online application of morphological rules. Lexical access is seen as a processing race between whole and composite forms; the winning form is the first to be activated. While there are several factors that may influence the outcome of the processing race, frequency of use¹¹ (i.e. a lexical chain’s resting level of activation in MOGUL terms) is regarded as central in determining the winner.

For example, the regular plural noun *pants* will be stored as both the whole chunk, *pants*, and as the compositional form, *pant*(N) + -s(plural). However, given the high frequency of the plural form and the relatively low frequency of the singular form (e.g. *a pant leg*), the term ‘pants’ is likely retrieved as a whole chunk. However, if the plural form has a low frequency (e.g. *fish* vs. *fishes*—meaning ‘species of fish’) then the plural form is more likely to be assembled online and thus accessed via its morphological components. While this example involves inflectional morphology, a similar case could be made for words involving derivational morphology, as in the case of *re-* in *reexamine* or *un-* in *unclassified*.

The question of lexical access is important to our account of intra-word codeswitching. In MOGUL, lexical elements may be stored as either whole chunks, compositional elements or both. When lexical forms are accessed, a processing race occurs between any available forms; the most active form wins selection. This notion of morphologically complex words being stored compositionally and formed via online

¹¹ This view is consistent with Jen Hay (2001), who argues for ‘relative frequency effects’ as opposed to ‘absolute frequency effects’ when discussing the decompositionality of lexical items.

computations is reminiscent of the view from DM that complex words are syntactically constructed from morpheme sized units stored in the lexicon. This point will be discussed in further detail in chapter 3 where elements of Distributed Morphology will be adopted to account for two morphemes from different languages combining to form an instance of intra-word codeswitching.

2.2.6. Functional vs. Lexical Categories

In MOGUL all linguistic elements (i.e. words, roots, & morphemes) may be classified into two groups—lexical categories or functional categories. Lexical elements traditionally consist of ‘open-class’ words—like nouns and verbs—as well as derivational morphemes which combine with lexical roots to form new words. SS&T (2014) claim lexical categories constitute the lexical store for a given language and are developed as (PS + SS + CS) representational chains in the regular manner discussed above.

Functional elements are more abstract and are present in the SS (i.e. syntactic features supplied by UG), but otherwise behave similarly to lexical elements. These functional categories are generally manifest in language production as ‘closed-class’ words—like determiners and pronouns—as well as inflectional morphology. In MOGUL, SS representations make up the heart of these functional categories, representing syntactic features like tense, number, agreement, etc.. However, just like lexical categories, functional categories may be associated with meaning (+ CS) and pronunciations (+ PS), albeit in a less than straightforward way. For example, the English plural morpheme *-s*, corresponds to the SS feature [+plural] and will be co-indexed with the CS representations meaning [+more than one] as well as a number of corresponding PS structures (i.e. [-s], [-z], [-əz]). The key point here, regarding our discussion, is that, despite ontological differences, lexical and functional categories behave in essentially the same manner.

2.2.7. The Nature of Processing in MOGUL

It is prudent to note that in the MOGUL framework processing is both incremental and dynamic. As noted by SS&T (2014), it is widely accepted as fact amongst cognitive scientists that language is processed in realtime, starting at the beginning of an utterance (i.e. left most side assuming an English orthography) and

moving rightward as the utterance is constructed—this is incremental processing. Additionally, processing in MOGUL is non-deterministic; this means that as MOGUL constructs a linguistic derivation (e.g. a sentence or phrase), the processors may respond to additional information in realtime which may cause the processors to move backwards to an earlier stage in the derivation and reconstruct representations accordingly—this is dynamic processing. Crucially, in MOGUL this means that language processing involves passing information back-and-forth between processors as opposed to linear serial models (Sharwood-Smith & Truscott, 2014) where information is only passed one way.

2.3.0. Acquisition by Processing Theory (APT) & States of Activation

To account for growth (i.e. language learning) within the MOGUL framework, SS&T appeal to their Acquisition by Processing Theory (APT) originally proposed in SS&T (2004). Truscott (2004) notes that APT originated as an attempted to formalize the Language Acquisition Device originally posited by Chomsky (1995) but ultimately evolved into an alternative account of learning and language development. The essence of their theory can be summed up in the following principle:

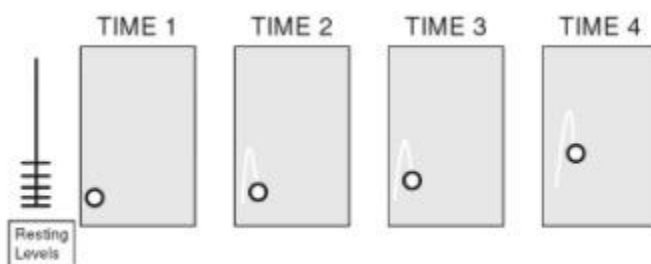
(x) Acquisition is the lingering effect of processing

In this theory, there is no specific device or mechanism that is responsible for the acquisition of novel linguistic elements; learning a new item is a byproduct of parsing novel input. When a processor establishes a novel representation (e.g. a novel bundle of CS features) it does so for the purpose of processing the current input (e.g. a newly encountered word) (p. 94). A representational chain is then formed from the most active content in each through co-indexation. The co-indexation of a new representation chain constitutes a lexical item which is then stored in long-term memory.

As stated above in section 2.1.7 (MOGUL & Lexicon), when a lexical entry is dormant it is stored in long-term memory (LTM). Dormant items in LTM maintain a resting level of activation which reflects the lexical entry's previous usage. The resting level of activation for a particular element is unique to that element. A high-frequency entry will have a high resting level of activation (e.g. *dog*, *hot*, *pizza*); low-frequency lexical entries (e.g. *troglydite*, *plutocracy*, and *demagogue*) will have a low resting level of activation and thus will be more difficult to recall.

MOGUL posits that each time a lexical entry becomes active, it is written in working memory and its activation level will spike before returning to a resting level slightly higher than before. This is illustrated in fig. 8 below, after each successive activation the item falls to a higher resting point of activation. Learning, in processing terms, is to increase an item's resting level of activation. Forgetting a word (i.e. language attrition) is the gradual decrease in an item's resting level.

Figure 8: Learning in APT



(SS&T, 2014: p. 95)

This can be exemplified in what is commonly referred to as the cocktail party effect—an individual's ability to pick out their own name even in a noisy, crowded room. In a MOGUL perspective, the countless times that an individual would be exposed to their own name will result in a high resting level of activation for the (PS+SS+CS) chain representing that individual's name. The byproduct of this elevated resting level of activation is the relative ease with which an individual's own name wins processing competitions.

To illustrate growth (i.e. learning) in the MOGUL framework, consider how a child might acquire irregular past tense forms (e.g. go → went)¹². As children are exposed to language they begin to establish PS representations which are associated to various SS and CS representations. As children have limited experience these chains initially have a low resting level of activation. However over time these resting levels of activation change to reflect the frequency of the intake; common intake increases its resting level while sparsely encountered or incorrect representations attrite. Over time, given exposure to linguistic intake, a child will develop a representational chain for the

¹² Language acquisition is individualistic; the process of acquiring irregular past tense forms as presented here is an example of one possible pathway that a child might take; it is not the only option.

word *go*. Assuming a regular developmental path, the child will also be exposed to a number of regular past tense forms marked with the regular past tense morpheme *-ed* (e.g. *watched, looked, painted, etc.*). Again, given time and exposure, the child will come to associate the *-ed* morpheme with a past action (c.f. Pinker & Prince 1988); the high frequency of the *-ed* morpheme will be reflected in a high resting level of activation. As such, when a child produces an utterance containing *go* in the past tense, the result may be the incorrect form *go-ed*; this is a common developmental error known as *over-generalization*. Overgeneralization occurs because the PS form ‘go’ is not co-indexed to the past tense SS representation; however, the PS representation for *-ed* is. This leads the PS to bundle the PS representations *go* and *-ed* together (i.e. *goed*) which are co-indexed to the SS and CS representation; the newly formed representational chain will then be stored in LTM.

However, this newly formed representational chain for the word *goed* is prescriptively incorrect and uncommon in adult speech. This means that children will receive no positive input (i.e. language exposure) to reinforce this form. Additionally, at this point in the child’s development, it can be assumed that they have encountered the past tense form *went* which will be co-indexed to a past-tense representation in the SS. The representational chains for *goed* and *went* will be identical in the CS and SS but differ in the PS and will be in competition. However, the relative low frequency of exposure (relative to *go* and *-ed*) to the form *went* means that *went* will have lower resting level of activation than the term *goed*. Over time, as a child is constantly exposed to language, the form *went* will be continuously encountered which will increase its resting level of activation. Eventually, as the representational chain for *goed* attrites due to lack of exposure, the form *went* will dominate every competition for selection; when this occurs the past-tense form *went* has been fully acquired.

In summation, in the MOGUL framework, there is no special mechanism (e.g. LAD) or trigger (e.g. error correction) for learning. The processor attempts to construct a representation for any input/intake it receives. If the processor has encountered the input before, a representation may be pulled from the information store (i.e. module-specific features stored in LTM). If the processor encounters a novel input/intake it may assemble a novel bundle of features (or use a partially preexisting set) as a representation

which is then written in the working memory for further processing (i.e. interfacing). After the input/intake has been processed, the activation level of the representation will fall down to a resting level of activation as the representation moves from working memory to long-term memory. The writing of new representations in LTM is how new items are learnt within a MOGUL processing framework.

2.4.0. MOGUL & Cognitive Context

A central element to language selection in MOGUL is cognitive context. The notion of cognitive context emanates from work on ‘mental models’ in the field of cognitive psychology (Johnson-Laird, 1980; Van Dijk, 1997) and may generally be characterized as a mental model an individual creates to reflect their environment, which is heavily influenced by personal experience (e.g. personal perceptions, pre-conceived opinions, etc.). Factors which may influence cognitive context include:

- Setting: location, timing of communicative event;
- Social circumstances: previous acts, social situation;
- Institutional environment;
- Overall goals of the (inter)action;
- Participants and their social and speaking roles;
- Current (situational) relations between participants;
- Global (non-situational) relations between participants;
- Group membership or categories of participants (e.g., gender, age).

(Van Dijk, 1997: p. 193)

Over the years, cognitive context has been used by a number of researchers to help account for various linguistic phenomena. For example, Van Dijk (1997) notes that Albrecht & O’Brien (1993) use mental models to explain the development of a referential base for pronominal anaphora when exploring local vs. global coherence. They were able to exploit mental models to account for a processing disadvantage when inconsistent antecedents were encountered in a long text (For instance, early in a text, Bill is described as old and weak; later he acts young and strong’ such inconsistencies hinder processing speed). Crucially, this highlights the involvement of context when matching antecedents

to their referents. This is illustrated below when looking at the meaning of the pronoun *their*;

- i) The students forgot their books.
- ii) The student forgot their book.

In sentence (i), *their* is used in its plural sense indicating possession by a group of students without reference to gender. The context surrounding sentence (i) could include a classroom with a number of textbooks left behind as well as the expectation that students should take their books with them. However, in sentence (ii), ‘their’ is used as a possessive gender-neutral third person singular pronoun. While this use of the pronoun *their* may be prescriptively considered ungrammatical, its use in sentence (ii) is the result of a mental model; for example, a situational context interacting with contemporary social values (i.e. gender-equality/gender assignment/LGBTI rights). The context for sentence (ii) will be almost identical to sentence (i); the only significant difference is the number of textbooks—only one book would have been left behind in example (ii). If the speaker knows the identity of the book’s owner, the owner’s identity (i.e. specifically, gender in this example) would become part of the speaker’s context and the speaker would default to using a possessive gender-specific third person pronoun like *his* or *her* to reflect the owner’s identity. However, if the speaker is unaware of the owner’s gender in this situation, cognitive context will generate the pronoun *their* to reflect that the owner’s gender is unknown.

When discussing codeswitching, cognitive context supplies the set of mental representations that lead an individual to produce one language instead of another in a given context. SS&T note that, “the activation of particular languages will... be triggered by given patterns of existing conceptual and affective structure: the initial source may be either internal or located in the observable environment” (SS&T, 2014: p. 198). In other words, language selection in production is cued not just by linguistic input/intake being processed in the mind/brain but by the social, political and environmental factors emanating from the real-world situation an individual is experiencing (i.e. external factors, such as street signs) as well as an individual’s perceptions of said experience (i.e. internal conditions, perceived language prestige).

2.4.1. Cognitive Context & Real-time Processing

MOGUL offers a processing perspective which takes place in realtime—stimuli are represented and processed in the order they are encountered as they are met. This means that cognitive context is constructed in realtime and subject to change as new stimuli are experienced. As a conscious person is constantly experiencing changes in their environment, cognitive context is constantly and continuously being updated to reflect the input/intake being processed at each moment. Essentially, context is fluid, constantly changing to reflect the experience a person is having.

2.4.2. Influences on Cognitive Context

While numerous factors can have an influence on context, three key factors have been identified in playing a central role in establishing a context which is conducive to language mixing. It is worth noting that this is not an exclusive list of contextual influences but rather a short-list of factors I propose to be central to establishing a context advantageous to language mixing. These factors are:

1. Linguistic Landscape
2. Identity of Language User/ Self-image
3. Identity of Interlocutor

An individual's linguistic landscape is a central external factor in establishing cognitive context (SS&T, 2014). A linguistic landscape refers to the real-world linguistic environment an individual finds them self in, taking into account visible aspects of language in a given environment, everything from road signs to gesture. This linguistic landscape is then mentally internalized and becomes part of cognitive context. For example, an individual witnessing a number of English street signs, as pictured below, would form an English linguistic landscape which would promote English productions.

Figure 9: Street signs contribute to a language user's linguistic landscape



It is important to note that a linguistic landscape is just as diverse and complex as any multicultural metropolis. While an individual living in a culturally homogenous city (such as Jinan City, China) may have a linguistic landscape dominated by a single language (in this case Mandarin Chinese), individuals living in cosmopolitan areas (culturally diverse), such as Hong Kong or Singapore, will develop intricate and multi-faceted linguistic landscapes to reflect the mosaic of languages they encounter (e.g. Mandarin, English, Cantonese, Malay, Indonesian, etc.). As such, more than one language can be present in a language user's linguistic landscape, which provides one of the contextual conditions to facilitate codeswitching in the bilingual.

Figure 10: Multilingual street signs contribute to a multilingual linguistic landscape



Another crucial factor is the language user's self-representation (i.e. a language user's view of themselves in their cognitive context). While, crucially, the language user must self-identify as a bilingual (or multilingual), self-representation also refers to a language user belonging to or identifying with various social/cultural groups (e.g. university students, musicians, beatniks, etc.) (Van Dijk, 1997). The identity (or perceived identity) of the interlocutor is equally important. The language user must know

that the interlocutor is a bilingual; additionally, if the language user and the interlocutor identify with similar social or cultural groups, the language user may perceive an affinity with the interlocutor which sets the stage for socially motivated codeswitching, spurred on by notions of solidarity and group belonging (Poplack, 1980).

An individual's preconceived perceptions of an interlocutor are also responsible for shaping cognitive context. For example, perceived ethnicity and cultural bias may arbitrarily associate skin colour with language (i.e. people from location X speak language Y). So, a Caucasian person living in Hong Kong—a former British colony—may be associated with English, while if the same individual were living in Haikou—a popular Russian tourist destination in China—they may be associated with the Russian language.

Crucially, cognitive contexts are unique to each individual. Individual people possess a wide array of personal preferences, natural talents, and personality traits, all of which may influence their perception of their experiences (e.g compare the hypothetical world-view of an artist vs. a soldier). As cognitive contexts are an amalgamation of personal experiences it is not only possible, but likely, that no two individuals will develop the same cognitive context despite exposure to the same stimuli.

2.4.3. Cognitive Context: An Example

A detailed hypothetical example of cognitive context can be found in Van Dijk (1997). To illustrate what a cognitive context may look like, Van Dijk speculates on the shape of a mental model (i.e. cognitive context) for Republican congressman, Mr. Rohrabacher during a congressional debate on the *Civil Rights and Women's Equity in Employment Act* in 1991. Van Dijk provides the following list of elements which combine to shape Mr. Rohrabacher's cognitive context during the debate:

1. Overall interaction and type of speech event: Congressional debate;
2. Location: The floor of the U.S. House of Representatives;
3. Date, time and timing: June 4, 1991, about 4:45 PM, yielded time of speaking (by debate organizer Mr. Hyde): 3 minutes.
4. Participant role: Current speaker;
5. Professional role: House Representative of California;
6. Affiliation: Republican Party;

7. Political ideology: Conservative;
8. General position on civil rights: Against extension;
9. Gender: Male;
10. Race: White;
11. Immediate opponent(s): Democrats;
12. General opponents: Liberals;
13. The social others: Minorities and women;
14. Current role: Speaker of intervention;
15. Hearers: The House;
16. Formal addressee: The Chairman of the House;
17. Overhearers: Public, voters, etc.;
18. Intention: Hold a good speech against bill;
19. Purpose: Defeat bill;
20. Overall goal: Defend business interests.

(Van Dijk, 1997: p. 215)

To begin we must note that this list is both hypothetical (Mr. Rohrabacher was not consulted) and non-exhaustive. To be clear, this example is Van Dijk's imagining of what Mr. R's cognitive context might look like. As it is impossible to literally peer into an individual's mind/brain, examples of mental models are, by their very nature, hypothetical; nonetheless they are efficacious in demonstrating how context can prompt the use of one linguistic form over another in a given situation. The list of elements combining to form a cognitive context is also non-exhaustive, in part due to the hypothetical nature of the example and in part due to space constraints. Nonetheless, this example is designed to highlight the complex assortment of elements that come together to form mental models.

The central idea behind cognitive context is that these mental models are used to guide language production. This is evident when looking at Mr. R's self-representation in his cognitive context and, consequentially, his choice of pronouns, as in "We just do not need any more laws in this area" (transcript line 6-7), and "We have got an underclass of people of all races trapped in poverty" (transcript line 24-25). Here, in a congressional debate, Mr. R. is not speaking as an individual but presents himself as speaking on behalf

of the American people; the *we* meaning “*we* in America” as opposed to the “royal *we*” or “*we* as in me”. The use of the pronoun *we* in this situation is a direct result of Mr. R’s cognitive context. That is to say that during such a debate Mr. R’s self-representation in his cognitive context would be dominated by notions (i.e. representations) such as “congressional representative”, “American citizen”, and “defender of the constitution” as opposed to less situationally relevant (albeit personally important) notions like ‘male’, ‘husband’, ‘practitioner of religion X’, etc.

2.4.4. How is Context Realized in MOGUL

In MOGUL, cognitive context may be viewed as an ‘internal mental model’ which is created, in part, via experience from the outside world. Consistent with the MOGUL representational and processing perspective, cognitive context is realized as the interaction between numerous tangentially active representations. In the model developed here, the CS is the locus of context, a central module to which numerous other modules, both linguistic and extra-linguistic, are interfaced. To illustrate a simplified example, context in the form of a physical environment is largely interpreted via POpS (i.e. Perceptual Output Structures). The most active features form a representational chain, while competing features will be slightly less active. However, these competing features are still active relative to their resting levels and contribute to developing a cognitive context. If a dog is present in the physical environment, POpS will construct a synchronized representation for the dog from the most active feature bundles in each perceptual module (i.e. Audio Structures, Visual Structures, etc.); however, competing representations, like CAT and WOLF, will have increased activation levels due to shared features in POpS which may represent numerous types of animals (e.g. shared VS features may include size, shape, colour, etc.). These secondary or tangential activations build a context around the representational chain formed from the most active features. In this manner, a representational chain and its surrounding context partially emanate from the same real-world experience¹³. This approach allows contexts to be accounted for using the basic MOGUL architecture and processing. Ultimately, the key takeaway

¹³ Innate properties in each module will also play a role in determining how stimuli are perceived.

from this discussion is that cognitive context may be realized as a natural byproduct of processing in MOGUL.

It should be noted that the complete story of cognitive context is much more complex than what has been presented above. All active modules contribute to developing a context; this will include all modules introduced thus far (e.g. POpS, AffS) as well as a great number of additional modules which account for various extra-linguistic aspects of cognition and are outside the scope of this paper (e.g. motor-control module, logic module, etc.). These modules are interfaced in myriad of ways which allow for changes in one module to have a cascading effect amongst co-indexed representations in other modules. To further complicate a complete account of context, experience and episodic memory will also play a central role by influencing how representations are co-indexed with each other. How various aspects of experience and episodic memory are realized in MOGUL is beyond the limits of this thesis; however, it would be possible to realize these phenomena within the standard MOGUL architecture and processing framework.

2.4.5. Cognitive Context, Conceptual Triggering & Language Selection

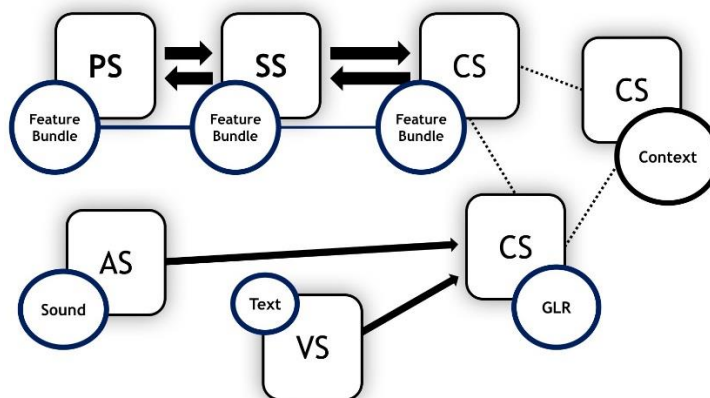
A fundamental question regarding bilingualism is concerned with a bilingual's ability to selectively speak one language or another without mixing or confusing elements cross-linguistically, or, as SS&T ask "How does the brain avoid babel"? This is a particularly pertinent question for researchers investigating codeswitching as before we can understand how two linguistic systems can be brought together, we must first examine how they are separated.

As such, any account of language mixing necessitates an account of language selection. Within the MOGUL framework, a rich cognitive context functions to activate elements in each module's information store that are associated with a specific situational context, as well as related associated representations which are stored in LTM. This process is known as *conceptual triggering* (SS&T, 2016). Conceptual triggering relies on a specific CS feature called a *general language representation* (GLR) (e.g. formal French, informal Mandarin, etc.) which increases the activation level of co-indexed

representations associated with a specific language variety¹⁴. SS&T state, “the contexts that have a triggering effect are particular perceptual and conceptual structures that are associated with elements of one of the person’s languages, including general language concepts like FRENCH or YORUBA” (2014, p. 199). Under this approach, language is viewed as a pattern of associations between representations in LTM and the cognitive contexts which activate them.

This account of language selection is motivated by the observation that bilinguals tend to use specific languages in specific contexts (e.g. language X is used at work but language Y is used at home) (SS&T, 2016). This association between context and language is realized as the co-indexation of contextual representations to a GLR. These general language representations are CS representations which are part of the complex CS representation and are co-indexed to feature bundles associated with the relevant language variety in the PS and SS; when a GLR from language X becomes active in the CS, all representations which are co-indexed to X-GLR receive an activation boost in addition to their current level of activation. It is worth noting that linguistic representations may be co-indexed to multiple GLRs, as in the case of cognates and interlingual homophones. This process is pictured below:

¹⁴ This terminology is a bit confusing. General Language Representations activate a specific variety of language but are referred to as ‘General’ because they ‘generally’ boost all representations associated with the specific variety (e.g. formal French), not a specific sub-set of lexical items/grammatical structures used to respond/interpret a specific utterance.

Figure 11: General Language Representations (GLRs)

(adapted from SS&T, 2016)

Legend : rounded squares = modules; circles = active features/feature bundles;**blue shapes = a specific language variety; black shapes = language neutral**

As seen above, in fig. 11, the active content in both the AS and the VS is co-indexed to a GLR-Blue in the CS. This GLR combines with other active features in the CS to form a complex CS representation. As the GLR-Blue becomes active in the CS, all co-indexed representations (i.e. blue representations) in neighboring modules will receive an activation boost; as such, feature bundles in the PS and SS associated to the GLR-Blue will have activation levels higher than feature bundles associated to other contexts which allows them to dominate any competition. In MOGUL this accounts for language selection. It is important to note that language selection, as it is discussed here, is a subconscious process. As such, it may be useful to imagine these processes as they apply to a bilingual toddler who may not be aware of which language they are producing¹⁵.

This discussion above implies that subconscious ‘language selection’ in MOGUL is really about establishing a cognitive context which correctly reflects the current experience of the mind/brain. This perspective allows us to equate context and language identity when discussing the mind/brain, as the GLR selection is triggered directly by cognitive context. Context is a cognitive construct but language is not—E-language¹⁶ is not recognized in the mind/brain during the processing of I-language. That is to say,

¹⁵ The same subconscious processes apply to both children and adults. However, toddlers lack metalinguistic knowledge (relative to adults) which serves to highlight the subconscious nature of language selection.

¹⁶ E-language, or Externalized language, is a collection of utterances which are observable in the real-world, as opposed to I-language which is Internal to the mind/brain and perceived only by an individual.

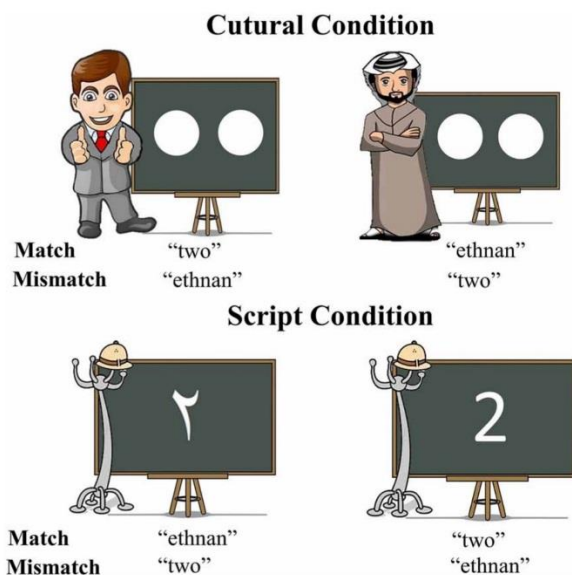
metaphorically, as linguists we have a bubble that we call *language*; inside this bubble we place all manner of linguistic phenomena— morphology, phonology, syntax, etc. MOGUL accepts that everything linguists place in the bubble is present in the mind/brain, but the bubble itself is not recognized. While MOGUL researchers use language names as labels for the context the GLR represents, there is no reason to believe that the brain does; what I call an English-GLR may be represented in the brain simply as being CONTEXT-A or CONTEXT-195—the label is arbitrary. This view is echoed by MacSwan (2005), who notes that there is no reason to expect language, as defined politically or culturally, to manifest in the mind/brain. This does not mean that we don't have the concept of 'language' (i.e. E-language) in our mind/brain; the notion of language exists as a social or political concept, as part of our meta-linguistic knowledge about language—the concept of 'English' in the MOGUL framework is meta-linguistic, not linguistic.

The process of language selection, as described above, accounts for how an individual is able to selectively boost activation levels of representations from a single language. However, this view is problematic for codeswitching researchers who ask: how does a bilingual form a single cogent string of linguistic elements from two different languages? As such, I will argue in this thesis to expand the role of cognitive context. In chapter 4, I will argue that real-time alternations to cognitive context are responsible for the activation of two language varieties in a simultaneous fashion. This will be important for explaining how a bilingual mode of communication can activate multiple languages via a single cognitive context which will be a central theme in chapter 4.

2.4.6. Evidence Supporting Cognitive Context

A recent MEG study by Blanco-Elorrieta & Pyllkanen (2015) has investigated the role of natural occurring language cues (i.e. ethnicity of the interlocutor & orthography) in triggering language selection. These language cues are partially responsible for establishing cognitive context. During their study, a number of Arabic-English speakers perform two number-naming tasks: a match task and a mismatch task. For each task there were two conditions, a script condition (i.e. orthographic) and a cultural condition (i.e. a culturally iconic picture). These conditions are displayed below.

Figure 12: Stimuli from Blanco-Elorrieta & Pylkkanen (2015: p.2)



During the match task, participants were shown an image and asked to name the number in the language indicated. The mismatch task required the opposite: participants were to name the number in the language which wasn't cued. The results of this study revealed that some language cues (i.e. orthography) are more effective at triggering a specific language than others. More specifically, the BE&P (2015) study reported, script—which is part of a linguistic landscape—was a much more effective cue than the cultural condition in terms of triggering language selection.

Neurologically, this distinction between the cultural and orthographic condition is demonstrated by a greater degree of activation in the ACC (Anterior Cingulate Cortex) during the cultural condition which is taken to be a result of greater processing difficulty. Prior research has implicated the ACC as having a major role in establishing attention (Abutalebi et al, 2013; Garbin et al, 2011; Abutalebi et al, 2008; Costa & Sebastian-Galles, 2014) which, in turn, is generally held responsible for performing the functions of alerting, orienting and executive attention/control (BE&P, 2015: p. 14).

Of particular interest is the notion of executive control, which BE&P describe as “the effort to retrieve... a word amongst competing responses” (2015, p. 14) during language production. As several studies have reported an increase in ACC activity for greater executive control demands (Garbin et al., 2011; Abutalebi et al., 2008; Costa &

Sebastian-Galles, 2014), BE&P suggest that the increase in ACC activity during the cultural condition reflects a greater effort to retrieve the target element; the cultural condition is a weak language cue, and requires additional cognitive effort to satisfy production goals. Alternatively, the lesser demand on the ACC during the script condition seems to indicate that script is a strong cue which appears to dominate cognitive context when selecting a language.

2.5.0. MOGUL & Executive Control

Traditionally, researchers have turned to the notion of executive control to help explain language control (Green, 1998; Abutelabi & Green 2015). Among its many functions, an executive control mechanism allows language users to suppress representations from language A while permitting representations from language B to surface. However, this view has been criticized by SS&T (2014) for serving as a homunculus (i.e. a cognitive mechanism which is responsible for making decisions). While this criticism conjures a comical image of a small person living inside one's head, flipping switches and pulling levers, it raises a very real question: if the homunculus controls language selection, who or what controls the homunculus? Additionally, early models of executive control proposed single mechanisms which were capable of performing a number of quasi-related functions which were otherwise unaccounted for. For example, Green discusses a control mechanism called the Supervisory Attentional System (see Green, 1998 for details), which, "...must command a variety of processes, including the construction or modification of existing schemas and the monitoring of their performance with respect to task goals" (Green, 1998: p. 69).

However, from a MOGUL perspective, where modules are believed to be 'expert systems' which perform specific tasks, there is no reason to believe that a single mechanism is responsible for performing these seemingly unrelated functions (i.e. goal maintenance, schema construction and conflict monitoring). MOGUL seeks to break down this executive control homunculus. "[There is] no single fixed executive control but [instead] different mental subsystems operate in a way that may be highly constrained" (SS&T 2014: p. 21). As such, the functions of executive control are broken down and attributed to specific modules which operate following the standard MOGUL representation and processing perspective. Of the many functions attributed to executive

control (e.g. goal maintenance; conflict monitoring; interference suppression; salient cue detection; selective response inhibition; task disengagement; task engagement; opportunistic planning), the notion of goal formation and maintenance is central to the account of intra-word codeswitching in MOGUL which will be discussed in chapter 4.

2.5.1. GOAL Representation in MOGUL

In MOGUL, goals are realized as goal representations which are constructed and interfaced to neighboring modules in a standard MOGUL fashion. SS&T note that there are two possibilities as to which goal representation emanate from; goal representation may form in the CS and be part of a complex CS representation; alternatively, goal representations may belong to their own goal module which functions like any other module in MOGUL. This paper will follow SS&T (2016) who adopt the former account where goal representations are manifest in the CS.

To illustrate this, let's consider how the notion of a 'goal', and consequentially 'goal maintenance', may be developed in MOGUL. SS&T (2016) claim that goal representations are a type of CS structure which help guide thought and action. At their most basic level, goal representations "serve the function of encouraging the satisfaction of basic needs" (SS&T, 2016: p. 5); these basic needs are non-linguistic and include the desire for food, water, bathroom, etc. These goals motivate action; goals with higher activation levels are prioritized and motivate an individual to perform tasks which satisfy the goal. For example, a basic goal like 'satisfy hunger' can be satisfied by eating.

As this paper is primarily concerned with language use, the goal representations which form to satisfy communicative and social functions are particularly pertinent. SS&T argue that these social goals (henceforth just goals) are formed from a set of primitive CS features representing social motivators like 'affiliation', 'power' or 'face'. These primitive goal features combine with complex CS representations and motivate a wide array of social behavior.

To illustrate goal representations, consider a business person attending a party with the goal of networking (i.e. creating new social and business contacts). As networking requires individuals to foster new relationships, the CS goal primitives representing 'increase affiliation' become active. Additionally, the motivation for networking is generally to create useful relations which can be exploited for business

purposes; this is represented in the CS with the feature bundle representing ‘increase power’ which signifies the business person’s desire to increase their prestige and status through social connections. Finally, as the business person needs to make a good impression, the CS feature bundle representing ‘maintain face/don’t look stupid’ becomes active to motivate socially acceptable behavior. These three CS features combine to form a complex CS goal representation which constitutes the goal ‘networking’ which will influence, among many other things, the style of language that the language user produces.

In MOGUL, the process of goal maintenance may be accounted for by combining this complex CS goal representation with value representations from the AffS. During language production, the goal representation interacts with POpS and AffS representations (i.e. contextual representations) as well as semantic CS representations (and all of their associated representations) which, in turn, affect the activation level of various lexical chains in the manner discussed above. The CS goal which is co-indexed to the highest value representation in the AffS will be the most active goal; the most active goal (i.e. the highest priority goal) is a central motivator which drives thought and action. As such, goals are maintained by their co-indexation to value representations in the AffS. Value representations are volatile and react in real-time to changes in cognitive context. For example, a speaker may be engaged in a conversation with the goal of ‘informing the listener about a specific event’. As the conversation continues, with no contextual changes, this goal is maintained by a high value AffS representation which motivates the speaker to continue the conversation. However, the sudden introduction of a loud noise, say a fire alarm, would lead to a shift in value representations which will have a cascading effect on the activation levels of various goals. When a speaker identifies the noise as a fire alarm (i.e. a warning of imminent danger), the goal of continuing the previous conversation will no longer be co-indexed to the highest value representation in the AffS, and the conversation will be abandoned due to a new goal of reacting to the fire alarm. Alternatively, if the speaker knew that the fire alarm was scheduled to sound, they may choose to ignore the fire alarm and carry on with their conversation; in such cases the speaker will have assigned a high value representation to

the goal of ‘informing the listener about a specific event’ and maintained this level of co-indexation despite contextual changes.

2.5.2. MOGUL and Communicative Modes

When discussing bilingualism, a number of language mixing researchers have suggested that bilinguals exploit ‘modes of communication’ (Grosjean, 2001; Poplack, 1980, 1988; MacSwan, 1999; Myers-Scotton, 1992; SS&T, 2014). ‘Communication modes’ refer to a specific configuration of mental apparatuses which accommodate a specific style of communication; this topic will be elaborated on in chapters 3 and 4. In MOGUL, these communicative modes are cognitive states where contexts and goals align to produce contextually relevant language. Of particular interest to this thesis is a language mixing or bilingual mode of communication. When a speaker engages in a bilingual mode, representations from two languages are equally active and compete on a level playing field. I will argue, in chapter 4, that this is made possible by an oscillating context which causes two languages to be conceptually triggered during the construction of a single lexical item. It is this configuration of contextual representations that allow the mixing of phonological systems during intra-word codeswitching.

The conditions which constitute a bilingual mode are a subset of the greater cognitive context. Within a MOGUL framework, communicative modes are realized as part of mental models. This means both external (e.g. linguistic landscape) and internal factors (e.g. goal representations) contribute to the subset of conditions which make up a bilingual context. In section 2.4.2, three key contextual factors (SS&T, 2014; Van Dijk, 1997) were previously identified which combine with goal representations to contribute to a bilingual mode of communication—these are conditions in the language user’s cognitive context which influence the activation levels of associated representations. Again, it is worth noting that this is not an exhaustive list of contextual influences (see Van Dijk, 1997), but rather factors that I have proposed as central to establishing a bilingual mode of communication.

Central factors which influence bilingual mode

1. Identity of Language User/ Self-image
2. Identity of Interlocutor
3. Linguistic Landscape

4. Social/Pragmatic/Linguistic Goal

The role of factors 1-3 was detailed above in section 2.4.2; these factors (i.e. self-image, interlocutor identify and linguistic landscape) are internalized representations which reflect a language user's experience in the real world. Crucially, these factors influence the configuration of a cognitive context. Additionally, goal representations will also influence context and will be central to establishing a bilingual language mode. While all necessary cognitive conditions may be in place to allow for codeswitching to occur, goal representations provide the language user with the motivation for constructing a codeswitch. While these motivations vary—from pride and pedantry to humor and solidarity—it is these goal representations which drive a language user to produce an intra-word codeswitch. This role of goal representation in intra-word codeswitching will be expanded upon in chapter 4.

2.6.0. Chapter Summary

This chapter introduced several cognitive structures and mechanism which will be exploited to account for intra-word codeswitching. The center piece of this chapter is MOGUL. This framework offers a real-time representational and processing perspective of language. Notably, MOGUL is a general model of cognition which incorporates a number of extra-linguistic processes which, nonetheless, have an influence on language production.

MOGUL exploits a modular architecture. Each module, regardless of whether it is linguistic or extra-linguistic, contains an information store, an integrated processor and an interface processor. Integrated processors construct representations (i.e. feature bundles) from module-specific features which are then interfaced to neighboring modules to form representational chains. The SS (syntactic module) and PS (phonological module) (i.e. the language core) construct linguistic representations which are co-indexed to the CS (conceptual module) to produce (PS + SS + CS) representational chains which constitutes a 'molecular view' of a word.

Growth within the MOGUL framework is accounted for via APT (Acquisition by Processing Theory). Representational chains are formed from co-indexing the most active feature bundles from each module. These representational chains have resting levels of activation based on past frequency of use. When a novel representational chain

is formed, it is written in LTM and has a low resting level of activation; as a representation is repeatedly activated, its resting level of activation increases and the representation will become more competitive and easier to recall. This change in activation levels is the central claim of APT which states that acquisition is the lingering effect of processing (SS&T 2004).

In MOGUL, language production relies on a number of extra-linguistic processes which interact with the language core. Central to this thesis, language selection emanates from cognitive context; contextual representations are co-indexed to GLRs (General Language Representations) which increase their activation levels accordingly. The most active GLR conceptually triggers all co-indexed linguistic representations, which provides these language/context-specific representations with an activation boost, thus allowing them to dominate any competition.

Additionally, language-orientated goal representations serve to motivate language use. Goal representations play a significant role in determining the shape and style of language production by increasing the activation level of linguistic representations that can satisfy the goal. These goal representations are co-indexed to value representations in the AffS; the stronger the value representation is, the more salient the goal, and the greater the increase in activation levels to linguistic representations that satisfy the goal.

The notions of cognitive context and goal representations come together to account for ‘modes of communication’. Contextual factors (i.e. linguistic landscape, bilingual interlocutor, etc.) and salient goal representations have a direct influence on the activation levels of linguistic representation. A mode of communication is a general configuration of contextual factors and linguistic goals which lead to the production of a specific style of language (e.g. monolingual mode, bilingual mode).

The key features of the cognitive framework, highlighted here, will play a vital role when accounting for word-internal language mixing—this is the central goal of chapter 4 where these topics will be revisited to account for intra-word codeswitching with the addition of a linguistic framework (i.e. chapter 3).

Chapter 3: The Linguistic Framework

3.0.0. Language Mixing

Language mixing is a general label that can be applied to a plethora of phenomena resulting from language contact. Of central concern to this thesis is the terminological distinction between *codeswitching* and *lexical borrowing* common amongst language mixing researchers (Poplack, 1980; Poplack, Sankoff & Miller, 1988; MacSwan, 2005; Myers-Scotton, 1992; Paradis & LaCharité, 1997; etc). For many researchers, the locus of this distinction is the degree of phonological and morphological integration the L2 element has into the L1—phonologically integrated words are classified as ‘borrowings’, while L2 lexical items which maintain L2 phonological characteristics are the result of code-switching. Additionally, codeswitched items are generally constructed using morphemes (i.e. Lx roots and Lx affixes) from a single language, while borrowed items take the morphology of the host language (i.e. Lx roots and Ly affixes).

Some researchers like Myers-Scotton (1992, 2000) claim that these factors are reflected cognitively with lexical borrowings being mentally integrated into the L1 as part of the lexicon while codeswitches maintain a separate L2 mental representation despite being inserted into an otherwise L1 sentence. The phonological differentiation between codeswitching and lexical borrowing is highlighted below in the Spanish-English example (3a) and (3b) taken from a bilingual (Spanish-English) Puerto Rican community. Additionally, examples (4a) and (4b) were collected from the Ottawa-Hull Corpus of French-English speakers and demonstrates morphological integration which is evident by the codeswitch (4a) maintaining its original affix while, in (4b), the borrowing adopts a French affix.

- | | | |
|---|---|-------------------|
| 3a) Leo un <u>magazine</u> [mægə ‘ziyn] | → | Codeswitching |
| ‘ I read a magazine. ‘ | | |
| 3b) Leo un <u>magazine</u> [maya ‘siŋ] | → | Lexical Borrowing |
| ‘ I read a magazine. ‘ | | |

(adapted from Poplack, 1980)

4a) Ils ont passé par la vitre arrière du char, pour voler le stéréo, et puis, speakers,
des affaires comme ça. → Codeswitching

'They climbed through the back window of the car, to steal the stereo, and
speakers, all that stuff.'

4b) Je serais pas capable de coper [kɔ'pe] avec. → Lexical Borrowing

'I couldn't cope with it.'

(adapted from Poplack, Sankoff & Miller 1988)

Significantly, any L2 items that are switched maintain their original L2 phonological and morphological form. This is illustrated, above, in (3a) where the word *magazine* is formed using an English lexical item—as opposed to the Spanish term for magazine which is *revista*—and phonology while the rest of the sentence is produced using Spanish grammar. Alternatively, (3b) is considered to be borrowed into Spanish as is evident by the presence of Spanish phonemes, namely [ɣ], in the English root. Crucially, this implies that the L2 item that is switched in (3a) will maintain its L2 mental representation during production, while (3b) is processed in the manner of an L1 lexical item. Likewise, (4a) illustrates that codeswitched items (i.e. *speakers*) carry their original morphology—as evident by the plural 's'—while borrowed items are morphologically integrated as is demonstrated in (4b) where the French suffix '-er'—pronounced as [e]—attaches to an English lexical item. It should be noted that, while a borrowed element is both phonologically and morphologically integrated into the host language, this integration is not always visible (e.g. 3b carries no affixes). These two criteria (i.e. phonological & morphological integration) will prove crucial to differentiating between codeswitching and borrowing.

3.1.0. Codeswitching & Borrowing

As noted above, the criteria for categorizing language mixing as a codeswitch or a lexical borrowing is at times unclear. Lexical borrowing generally refers to a single L2 lexical element (e.g. a word or idiomatic phrase) that fits in a syntactically L1 sentence. Crucially, lexical borrowing involves inserting an L2 lexical item into an L1 frame where

the L2 lexical item conforms to L1 morphology and phonology. This is highlighted above in (3b) where the English term *magazine* is phonologically integrated into Spanish. In their investigation into loanword adaptation Paradis & LaCharité (1997) provide the following definition of a lexical borrowing (i.e. loanword):

Loanword (Lexical Borrowing): An individual L2 word, or compound functioning as a single word, which:

- (a) is incorporated into the discourse of L1, the recipient language;
- (b) has a mental representation in L1 (as opposed to codeswitches—Myers-Scotton, 1992) and thus;
- (c) is made to conform with the phonological constraints of L1

(Paradis & LaCharité, 1997, adapted from Poplack, 1988)

Importantly, these criteria highlight the fully integrated nature of lexical borrowings. In order for a lexical item to be considered a loanword it must appear in the host language (as in (a)). As loanwords are used by both monolinguals and bilinguals, they must be incorporated into the grammar of the host language¹⁷ (as in (b)). Finally, part of being incorporated into the host language means conforming to the phonological norms (as in (c)).

Poplack, Sankoff & Miller (1988) note that lexical borrowings may appear in two forms; (1) *established borrowings*—also called loanwords; these L2 lexical items are fully integrated into the L1 and are commonly accepted amongst both the bilingual and monolingual speech community—and (2) *nonce borrowings*—integrated L2 lexical items that have a very low frequency (possibly just a single occurrence) and may not be judged as grammatical by the entire language community. For the purpose of this thesis, it is important to note that, while the monolingual community will make use of established borrowings, nonce borrowings are solely the product of bilingual speakers (Paradis & LaCharité 1997). As noted by Myers-Scotton (1992), these nonce borrowings can show a high degree of morphological integration as exemplified below in an English/Tamil mix:

¹⁷ Note that exceptions exist. When a large number of loanwords enter a language as a group—usually motivated by a historical event (i.e. Norman invasion)—they may maintain some of their original features (i.e. irregular plurals like *attorneys general*); this is the result of culture and politics not cognition (i.e. French was the language of law/administration in Norman England) (McFeteridge, 2008).

- 5) anta **car**-ei drive paNNanum
 That car-ACC drive do.must
 We must drive that car.

(Sankoff et al., 1990: p. 80)

The notion of lexical frequency combined with the degree of morphological and phonological integration for mixed language elements creates an identification problem for researchers attempting to classify language mixing into a codeswitch/borrowing dichotomy. Authors, like MacSwan (2005) and Poplack (1980), tenaciously argue that codeswitching and lexical borrowing are distinct phenomena (presumably with distinct cognitive processes and mechanisms); ambiguous cases like the intra-word codeswitching (e.g. *den field-a*) and example (5) are treated as nonce borrowings. However, dissenting authors, such as Myers-Scotton, argue that codeswitching and borrowing exist on a sort of continuum, with codeswitches potentially representing the initial state of a loanword which becomes increasingly integrated into the borrowing language. It is precisely these elements that fall between the traditional classification of ‘codeswitch’ and ‘lexical borrowing’ that we are interested in. I will argue that some instances of nonce borrowings (i.e. examples (2) & (4)) should be classified as examples of intra-word codeswitching—this will be evident in chapter 4. The exact relationship between codeswitching and lexical borrowing is of great interest to this thesis and will be explored in greater detail in chapter 4.

3.1.1. Codeswitching and Speech Mode

Codeswitching is often viewed as a speech style in which (fluent) bilinguals switch between two languages (MacSwan, 2005; Poplack, 1980). Codeswitchers may substitute L2 words (lexical codeswitching), phrases (intra-sentential codeswitching) or entire sentences (inter-sentential codeswitching) during the course of a single conversation (often with a single interlocutor) (Poplack, 1980). Of particular interest to this paper is the notion of intra-word codeswitching where language users construct a word with morphemes from two languages. Examples of each type of codeswitch appear below in Table 1.

Table 1: Types of Codeswitching

Type	Example	Languages	Sources
Sentence Level	Ni hao. Where are you going? Hello/How are you? Where are you going?	Mandarin/ English	N/A
Phrase Level	<u>No tienen ni tiempo</u> sometimes for their own kids, and you know who I'm talking about. <u>They don't even have time</u> sometimes for their own kids, and you know who I'm talking about.	Spanish/ English	Poplack (1980)
Word Level	Me iban a <u>layoff</u> They were going to <u>lay me off</u> .	Spanish/ English	Poplack (1980)
Intra-word	den <u>field</u> -a that <u>field</u>	Norwegian/ English	Alexaidou et al. (2015)

When an individual language mixes they may do so for a number of different reasons. For many bilinguals, codeswitching emanates from the social-cultural implications that follow from the use of a particular language than the desire to improve the clarity of a message. Codeswitching in such a manner may mark membership in a particular social class or group; be used as a sign of solidarity; serve a conversational function such as drawing attention to a specific topic or humor; etc. Codeswitching of this variety has less to do with linguistic processing and more to do with executive control and goal orientation—“codeswitching of this type is essentially the same phenomenon as control of action” (SS&T, 2014: p. 203).

Additionally, speakers may choose to switch from one language to another as a way to help and accommodate the listener(s); switching of this sort not only makes it easier to express a specific idea but may actually improve the accuracy of the message. That is to say that a speaker may switch from language A to language B when doing so will allow the listener to better understand the desired message. As SS&T point out, “Improving communication is a standard function of code-switching” (SS&T, 2014: p. 200).

Early work by Poplack (1980) probed the social conditions surrounding code-switching. Some researchers (Haugen, 1950) hypothesized that code-switching occurs in situations where speakers are attempting to fill a lexical need in their L1 with elements from their L2—this is sometimes referred to as *filling lexical gaps*. While Poplack (1980) and Poplack, Sankoff & Miller (1988) report that some instances of codeswitching seem consistent with Haugen’s claim (notably the use of specialized nomenclature as in English logging terms borrowed by French lumberjacks: *drave* ‘lumber drive’; *slab* ‘firewood’), the vast majority of switches seem without social or functional motivation. This observation becomes even more evident when Poplack (1980) discusses examples of codeswitching where the sole element switched is a functional category like a determiner or object pronoun. This is exemplified in the example below where the noun and adjective are Norwegian but the determiner is English:

- 6) **the** eneste ung-en
 the only child-DEF.M
 ‘the only child’

(Alexaidou et al., 2015)

Clearly, in example 6, the English and the Norwegian definitive determiner could serve the same function. Ultimately, Poplack (1980) concludes that “there is no good reason” (p. 614) for these types of codeswitches—codeswitching of this type does not clarify or improve the message being communicated in any conceivable way. Thus, to account for what appears to be nearly unconstrained code-switching (i.e. one can codeswitch anywhere in a linguistic structure; MacSwan 2005), Poplack (1980), in a view shared by MacSwan (1999, 2005), suggests that code-switching may result from a mixed-speech mode for bilinguals. That is to say in specific contexts, bilinguals may configure their cognitive apparatus in such a way that language mixing is the natural result of language processing. Following Grosjean (2001) this thesis will use the term *bilingual mode* (as opposed to *mixed speech mode* or *codeswitching mode*). In such circumstances, it is the act of switching, not the location of switch, that may carry meaning—codeswitching promotes solidarity; the type of syntactic elements switched are not relevant to fostering solidarity. Metaphorically, codeswitching of this sort functions as an esoteric verbal wink. The linguistically unmotivated language mixing, seen in example

(6) may be the result of such circumstances. This will be explored in greater detail in chapter 4, where I will argue that in the MOGUL framework this *bilingual speech mode* is engaged as result of two languages being conceptually triggered simultaneously which allows intra-word codeswitching to occur.

3.2.0: Generative Grammar & Language Mixing

3.2.1: Constraints on Codeswitching

There has been a long history of researchers attempting to establish syntactic boundaries on codeswitching—a task which has produced little agreement. For example, Belanzi, Rubin & Toribio (1994, cited in MacSwan, 2005) postulated a Functional Head Constraint where a codeswitch may not occur between a functional head and its complement. This analysis relies on the presence of a ‘language feature’ or tag, such as [+English] or [+Spanish]. The constraint is formulated below:

The Functional Head Constraint

The language feature of the complement f-selected by a functional head, like all other relevant features, must match the corresponding feature of that functional head.

Notably, this constraint prohibits codeswitching from occurring between a functional head and its complement (i.e. [modal + VP] or [Demonstrative + NP]). This is because the language feature on the switched element must match the language feature on the functional head. However, in this account, as the functional head constraint only applies to f-selected structures (i.e. structures where a complement is selected by a functional head), switching between lexical heads and their complements is permissible. However, there are dozens upon dozens of empirical examples of codeswitching which blatantly violate the Functional Head Constraint. Previously introduced examples like *den field-a*, and examples (5) & (6) are all counter-examples of this constraint. An additional example is provided by MacSwan (2005), seen below, where the Nahuatl indefinite article *se* introduces the Spanish noun *hombre*.

- (7) Se hombre kikoas se kalli.
 se hombre 0-ki-koa-s se kalli
 a man 3S-30s-buy-FUT a house
 A man will buy a house.

The key take away from this discussion is that previous efforts to place grammatical constraints on codeswitching have proven largely problematic. Table 2, presented below, is a summary of findings based on codeswitching corpora taken from MacSwan (2005); the sheer lack of consensus on proposed descriptive boundaries for code-switching is noteworthy and has hampered the advancement of any one particular account of codeswitching. This is largely problematic and exemplifies the need for a unifying framework, such as MOGUL.

Table 2: Review of Proposed Constraints on Codeswitching

<i>Item ref #</i>	<i>Descriptive boundaries (+ = code switch)</i>	<i>Reported in ...</i>	<i>in disagreement with ...</i>
1a	<i>because + CP</i>	Gumperz 1976	Poplack 1981 Sankoff and Poplack 1981 Mahootian 1993
1b	<i>conj + CP</i>	Gumperz 1976	Poplack 1977 McClure 1981
2	<i>that + IP</i>	Belazi, Rubin and Toribio 1994	Bentahila and Davies 1983 Mahootian 1993
3a	<i>have + VP</i>	Belazi, Rubin and Toribio 1994	Di Sciullo, Muysken and Singh 1986
3b	<i>modal + VP</i>	Belazi, Rubin and Toribio 1994	Di Sciullo, Muysken and Singh 1986
3c	<i>to + V</i>	Timm 1975	Lipski 1978 Poplack 1981 McClure 1981
3d	<i>Aux + V</i>	Timm 1975	Lipski 1978 Poplack 1981 McClure 1981 Mahootian 1993
3e	<i>Neg + V</i>	Timm 1975	
4a	<i>Q + NP</i>	Belazi, Rubin and Toribio 1994	Bentahila and Davies 1992 Mahootian 1993
4b	<i>Demonstrative + NP</i>	Belazi, Rubin and Toribio 1994	Nishimura 1985 Bentahila and Davies 1992 Mahootian 1993
4c	<i>Article + NP</i>	Belazi, Rubin and Toribio 1994	Brown 1986 Bentahila and Davies 1992 Mahootian 1993
5a	<i>N + Adj Adj from Adj-N language, N from N-Adj language</i>	Gumperz 1976 Lipski 1978 Belazi, Rubin and Toribio 1994	Bokamba 1989 Mahootian and Santorini 1996
5b	<i>Adj + N Adj from N-Adj language, N from Adj-N language</i>	Belazi, Rubin and Toribio 1994	Poplack 1981
6a	<i>Subject pronoun + V</i>	Timm 1975 Gumperz 1976 Lipski 1978	Poplack 1981 Woolford 1983 Nortier 1990 Eid 1992 Bentahila and Davies 1983
6b	<i>V + object pronoun</i>	Timm 1975 Gumperz 1976 Lipski 1978	Poplack 1981 Mahootian 1993
6c	<i>clitic + V or V + clitic</i>	Timm 1975	undisputed
7	<i>A switch involving a bound morpheme</i>	Poplack 1981 Sankoff and Poplack 1981	Nishimura 1985 Mahootian 1993 Myers-Scotton 1993

(MacSwan, 2005: p. 66)

As demonstrated above, in Table 2, attempts to place syntactic boundaries on code-switching have been largely unsuccessful. This lack of empirically consistent and universally accepted boundaries on codeswitching, combined with Minimalist theoretical assumptions (i.e. a null theory) has motivated MacSwan's central assertion regarding code-switching; stated below:

MacSwan's Assertion:

- i) Nothing constrains code-switching apart from the requirements of the grammars involved

As MacSwan is quick to point out, this does not mean language mixing is subject to *laissez faire* style codeswitching, but rather that boundaries on code-switching are established via the union of grammars between language-X and language-Y, and not by some aspect of the human computational system. Following MacSwan, this may be formally stated as $\{G_x \cup G_y\}$ where G_x is a grammar of L_x and G_y is a grammar of L_y . As noted above, language specific variation in codeswitching boundaries is attributed to parametric variation in the lexicon (Borer & Wexler, 1987); any natural codeswitching boundaries that arise are the product of a union between the grammars of L_x and L_y and unique to the specific blend of languages in which they appear.

MacSwan's minimalist approach to codeswitching is generally referred to as a 'null theory' in codeswitching literature (MacSwan, 2005; Grimstad, Lohndal & Afarli, 2014; Alexaidou, et al. 2015). While it may seem odd to refer to standard MP assumptions as a 'null theory', the term is used in response to alternative theories of codeswitching—notably, the Matrix Language Model of codeswitching (Myers-Scotton, 1993, 2001)—which argue that codeswitching is the result of processes or mechanisms unique to language mixing. As such any account of codeswitching which does not posit extra machinery is considered a 'null theory' in codeswitching literature regardless of whether it follows MP assumptions or not.

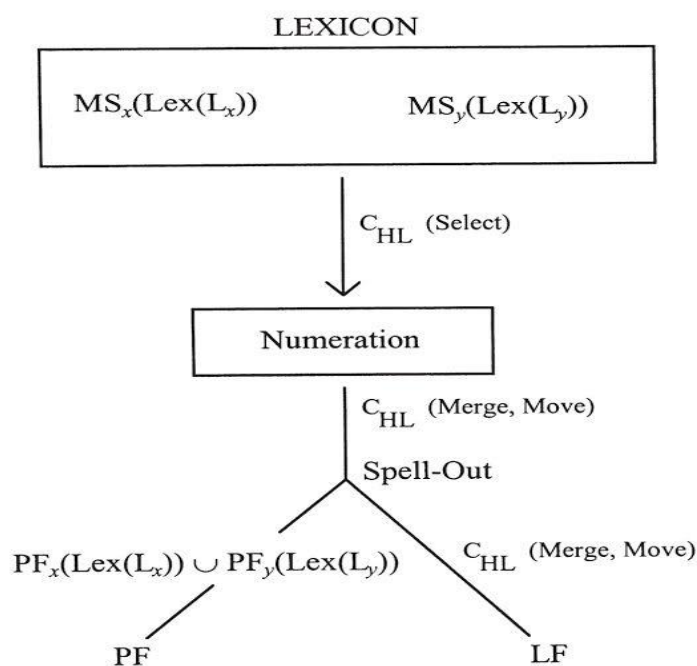
3.2.2: MacSwan, MP and the Language Faculty

MacSwan (2005, 2014) developed his model of codeswitching within the Minimalist Program (Chomsky, 1995) and generally adheres to lexicalist assumptions regarding the mental organization of the lexicon. Working within Minimalism, MacSwan sought to develop a 'null theory' of codeswitching; that is to say a theory of codeswitching which does not rely on a special 'codeswitching' mechanism. Under this view, both monolinguals and multilinguals possess the same linguistic computational system; language is differentiated in the lexicon. Each language is said to have its own lexicon which is part of its own Memory Store (MS). Elements stored in the lexicon contain syntactic, phonological and semantic features, which are projected from the

lexicon to the rest of the derivation (i.e. The Lexical Integrity Principle, Bresnan & Mchombo, 1995). MacSwan (2005) provides the following sketch of the bilingual human language faculty in the Minimalist Program (MP).

I will now provide a brief overview of the language faculty in MP (for a more comprehensive review see Chomsky, 1995; MacSwan, 2005) as is relevant to our story. As seen in fig. 13, below, there are two key components that constitute the language faculty: the lexicon (Lex) and the Computational System for Human Language (C_{HL}). The C_{HL} is responsible for generating linguistic structures. While the C_{HL} is stable and invariant across all human language, the lexicon varies from language to language (i.e. Lex(L_x) to Lex(L_y)) and is held responsible for cross-linguistic variation in language; the lexicon is the source of parametric variation (in the sense of Principles and Parameters; Chomsky & Lasnik, 1993) in languages. As illustrated below in fig. 13, in the case of a bilingual, MacSwan argues that lexical items from separate languages are mentally compartmentalized into two language groups yielding the appearance of two sub-lexicons.

In MP, phrase structure is derived from the lexicon. The C_{HL} uses the operation *Select* to pull lexical items from the lexicon to form a *numeration* (i.e. lexical array). In the case of codeswitching, elements may be selected from both Lex_{L_x} and Lex_{L_y} (i.e. sub-lexicons). After the numeration is formed, the C_{HL} then applies the operations *Merge* which takes selected lexical items in numeration and combines them to form new, hierarchically arranged syntactic object; and *Move* forms new structures by means of probe starts at the top-edge of the structure, selects an element from inside the structure and re-locates the element to become a specifier of the probe. These movements are driven by feature checking; movement may be overt (driven by strong features) or covert (driven by weak features). Overt movements are visible at both Phonological Form (PF) and Logical Form (LF), covert movements are only visible at LF.

Figure 13: A Bilingual Minimalist Grammar

(MacSwan, 2005: p. 75)

MacSwan's model enjoys a broad range of empirical support from researchers investigating intra-sentential codeswitching. A number of empirical studies (Poplack, 1980; Poplack & Sankoff, 1988) lend support to MacSwan's lexicalist approach to codeswitching. However, recently, a number of researchers interested in language mixing at the word level (Lohndal, 2013; Grimstad, Lohndal & Afarli, 2014; Alexiadou et al., 2015; Gonzalez-Vilbazo & Lopez, 2011) have noted that MacSwan's lexicalist assumptions (i.e. the inherent link between a syntactic root and its phonological exponents) mean this model makes it difficult to account for intra-word codeswitching.

3.2.3. Intra-word Codeswitching and the PF Disjunction Theorem

Part of the natural fallout of MacSwan's Lexicalist approach to codeswitching is a ban on intra-word codeswitching. This claim is supported by observations made by Poplack (1980) who noted a prohibition on intra-word codeswitches. Gonzalez-Vilbazo & Lopez (2011) formally describe this type of code-switch as $[w M_{L1} + M_{L2}]$, where 'w' stand for a word and 'M' for a morpheme; crucially, each M maintains its original phonology. Poplack (1980) illustrates this illicit type of codeswitch in her classic examples below, all of which were unattested in her language mixing corpus:

- 8a) *Juan esta eat-iendo
 Juan be/1Ss eat-DUR
 Juan is eating.
- 8b) *Juan eat-o
 Juan eat-PAST/3Ss
 Juan ate.
- 8c) *Juan com-ed
 Juan eat-PAST
 Juan ate.
- 8d) *Juan eat-ara
 Juan be/1Ss eat-FUT/3Ss
 Juan will eat.

(Poplack, 1980)

MacSwan argues that an MP framework can account for this ban on intraword codeswitching. As stated by Chomsky (1995), word formation occurs before the numeration is formed. MacSwan (2005) goes on to assume that each lexical item selected for the numeration will trigger a language specific grammatical system for encoding. As code-switching invokes the union of G_x and G_y , the computational system (i.e. C_{HL}) must allow the merger of these two grammars. This does not appear to be a problem for syntactic features; as MacSwan states, “each lexical item imposes certain requirements on the derivation in terms of the encoded features, and syntactic operations need take no notice of what particular language a lexical item is associated with” (2005: p. 72), thus, “code-switching in syntax appears to be constrained only by the operations of C_{HL} on lexically-encoded features” (2005: p. 72).

However, phonological operations are hierarchically ordered and the hierarchical order of phonological operations differs from language to language (MacSwan cites Bromberger & Halle, 1989). As such, according to MacSwan, the union of PF_{L_x} and

PF_{Ly} is not possible because the order of phonological operations is not preserved under union. Ultimately this entails that intra-word code switching is not possible. This is stated formally in the **PF Disjunction Theorem** below:

PF Disjunction Theorem

- i) The PF component consists of rules/constraints which must be (partially) ordered/ranked with respect to each other, and these orders vary cross-linguistically
- ii) Codeswitching entails the union of at least two (lexically encoded) grammars
- iii) Ordering relations are not preserved under union
- iv) Therefore, code-switching within a PF component is not possible

(MacSwan, 2005: p. 73)

Crucially, the PF Disjunction theorem bans the switching of phonological systems between grammatical morphemes. As such, codeswitching of the type [$w M_{Lx} + M_{Ly}$] is prohibited as the ordering of Lx phonological operations cannot be unified or merged with the ordering of Ly phonological operations during the PF spellout of a single word. Note that this is distinct from intra-sentential codeswitching where one set of phonological rules is applied to part of the sentence and a second set of phonological rules is applied to the codeswitched constituent—codeswitching at the constituent level does not rely on the merger of systems but rather an alternation between them¹⁸ (e.g. a switch from system A to system B).

Consider the illicit Spanish/English codeswitch *eat-iendo*, as seen above in example 8a. The selection of the English lexical item *eat* for the numeration will trigger the application of an ordered set of English phonological rules during PF spellout. Any attempt to add the Spanish morpheme *-iendo* to the derivation will result in the attempted application of an additional set of phonological rules (i.e. Spanish phonology). As the type and order of phonological rules regarding phonemic processes (e.g. stress assignment, phonological interactions, etc.) differ in the two languages, the C_{HL} has no

¹⁸ It is unclear how phonological processes which apply to the sentence level (i.e. sentence stress) are dealt with in MacSwan's model. MacSwan's PF Disjunction Theorem appears to be restricted to the word level.

way of re-ordering the two systems while maintaining any kind of coherent ordering. Thus, according to MacSwan, any attempt to merge phonological systems during PF spellout will result in an unpronounceable output and the derivation will ultimately crash. However, this analysis becomes largely problematic when applied to empirical examples of intra-word codeswitching found in Alexaidou et al. (2015); intra-word codeswitching does exist but is a clear violation of MacSwan's PF Disjunction Theorem.

3.3.0. A Brief review of Distributed Morphology and Language Mixing

A solution to the theoretical quandary found in MacSwan's lexicalist account of language mixing can be found in the non-lexicalist assumptions of Distributed Morphology (DM). Recent work by a number of researchers (Marantz, 1997; Embick & Noyer, 2007; Lohndal, 2013; Grimstad, Lohndal, & Afarli 2014; Arad, 2003; Harley, 2014) has proposed an alternative account of how lexical roots and morphemes operate in the grammar. Of particular relevance to this paper, DM offers a model where under-specified morphological elements compete for late-insertion into a fully generated syntactic frame, complete with syntactic terminal nodes.

3.3.1. Distributed Morphology: The Syntactic Frame

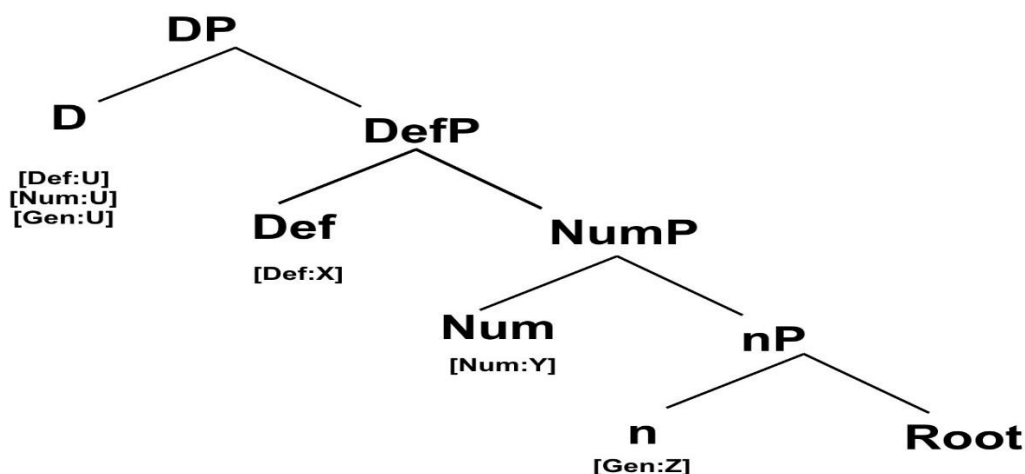
Under this approach words are derived from syntactic structures (i.e. *exo-skeletal* frames¹⁹). Grimstad, Lohndal, & Afarli (2014) note that there are two possible options to account for the generation of syntactic frames: they may be generated by functional features/feature bundles, or the operation Merge may operate freely. The exact nature of how these frames are generated is not vital to this paper. However, following Grimstad, Lohndal, & Afarli (2014), this paper will assume the former where "the abstract building blocks of syntactic structures are functional features and functional feature matrices... we will assume that the functional elements instantiate feature matrices, whereas lexical content items are freely inserted into designated lexical slots" (p. 221)

A general example of an *exo-skeletal* template is found in Grimstad, Lohndal and Afarli (2014), featured below in fig. 14. Their analysis follows work by Julien (2009)

¹⁹ Grimstad, Lohndal & Afarli borrow the term *exo-skeletal* from Borer who argues for an augmented version of DM. While I am presenting a 'standard version' of DM in this section, I will follow Grimstad, Lohndal & Afarli in using the term synonymously with *syntactic frame*.

who argued for the inclusion of a DefP to account for double definiteness marking. While Grimstad, Lohndal & Afarli make use of several standard functional heads (i.e. D – determiner; Num – number; Gen – gender; n – nominalizing head), they follow Julien (2009) in arguing for a functional DefP; this unusual analysis is meant to account for double definiteness marking in Norwegian. Norwegian nouns are inflected with both number and definiteness and classified by gender. Like English, Norwegian nouns are introduced by determiners; unlike English, Norwegian determiners must agree with the number and definiteness of the noun. This agreement between a noun and its determiner is known as ‘double definiteness’ marking; this phenomenon is illustrated in the intra-word codeswitch *den field-a*, where the Norwegian determiner *den* agrees with the Norwegian suffix *-a* in both number and definiteness. Norwegian double definiteness marking will be revisited in greater detail below in section 3.3.4.

Figure 14: Norwegian Syntactic Frame



(Grimstad, Lohndal & Afarli, 2014: p. 224)

This syntactic frame is constructed using abstract feature bundles (i.e. feature matrices [FM]). The result is a syntactic template with specified syntactic features but no phonological or semantic content. The phonological items of a word which matches the abstract features in the template can then be inserted into the derivation. This process, known as *vocabulary insertion*, will be expanded upon below in subsequent sections.

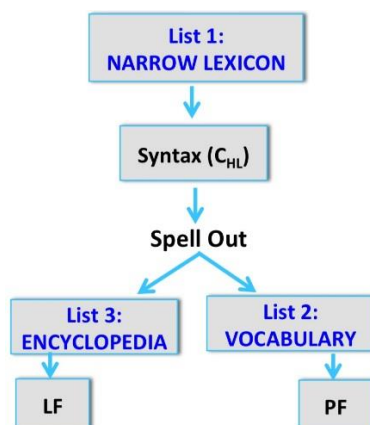
In a DM account of language mixing, the notion of lexical decomposition also plays a central role in allowing intra-word codeswitching to occur. Lexical

decomposition is the notion that category-neutral lexical roots (e.g. *cat*, *man*, etc.) combine with one or more functional heads in the syntax (e.g. *Gen*, *Def*, *Num*, etc.); proponents of lexical decomposition argue that this accounts for complex syntactic meanings (Halle & Marantz, 1994). Notably, in terms of language mixing, it is the construction of syntactically complex words, where a root is from language X and the syntactic features are from language Y, which ultimately allows intra-word codeswitching to occur in DM. This is in contrast to MacSwan's Lexicalist approach, in which morphologically complex words are viewed as syntactic atoms which cannot be syntactically decomposed, thus preventing intra-word codeswitching.

In DM, the construction of words and phrases is driven by a single engine, which means that words and phrases are expected to behave in the same fashion (i.e. the same syntactic principles, domains and mechanisms are at work cross-linguistically). If codeswitching is allowed at the sentence or phrase level, then there is an expectation that it will also occur at the word level. This notion is echoed by Grimstad, Lohndal & Afarli (2014) who argue that intra-word codeswitching is an expected consequence of a DM account of language mixing.

3.3.2. Distributed Morphology: The Lists

Notably, in Distributed Morphology the content of the lexicon is said to be distributed over three separate lists: Narrow Lexicon, Vocabulary, & Encyclopedia. This is illustrated below in fig 15. List 1, the narrow lexicon, which may also be referred to as the *syntactic terminal nodes*, contains "the atomic roots of language and the atomic bundles of grammatical features" (Marantz, 1997: p. 203). The second list, the Vocabulary, contains phonological information which is inserted into the syntactic terminal nodes post-syntactically; this accounts for the phonological realization of lexical items. Foreshadowing their account of intra-word codeswitching, Grimstad, Lohndal & Afarli (2014) note that this type of language mixing is a product of the application of list 2, where vocabulary items from two different languages are inserted into a mono-lingual syntactic frame. Once the derivation has phonological form, the morphosyntax is interpreted by the intentional/conceptual system and the root morpheme, which carries 'special meaning', is interpreted by the encyclopedia (i.e. list 3). These lists will be expanded upon below.

Figure 15: Stages of the Derivation

(Burkholder, Mathieu, & Sabourin, 2017)

Crucially, this approach distributes morphology across the lexicon (i.e. the 3 lists). Grimstad, Lohndal & Afarli (2014) note, “If features have a pragmatic, semantic or phonological basis, one could argue that rather than syntacticizing such features, the relevant effects should be analyzed in these components in order to avoid duplicating the analysis across grammatical components” (p 221). By separating the syntactic, phonological and semantic components of a word, DM allows for a molecular view of a word which is, certainly, compatible with the MOGUL account (see section 2.2.0. for details).

The syntactic terminal list contains two types of primitives; category neutral lexical roots and abstract morphemes (Grimstad, Lohndal & Afarli, 2014). Lexical roots, (i.e. $\sqrt{\text{TABLE}}$, $\sqrt{\text{CAT}}$, $\sqrt{\text{RUN}}$, etc.) reside in the memory store. While, the exact nature of roots in DM is still a topic of debate (Harley, 2014), this paper will follow Grimstad, Lohndal & Afarli (2014) in assuming that roots contain no grammatical features themselves and are underspecified both phonologically and semantically. Multi-morpheme words are formed by merging roots with abstract morphemes (e.g. roots combine with the nominalizing head - *n* - to form nouns (Burkholder, 2017)). Significantly, all root vocabulary items are part of the same memory store and are not partitioned by language identity or membership. (Meta) knowledge regarding language identity is part of the Encyclopedia (Grimstad, Lohndal & Afarli, 2014).

However, abstract morphemes, also known as features or feature bundles, are part of language specific sub-lists (e.g. List 1: Sublist A) (Grimstad, Lohndal & Afarli, 2014).

The features that constitute these abstract morphemes, regardless of language membership, are taken from the same store—the pool of syntactic primitive features—but are organized into a sublist of language specific feature combinations or bundles. This allows for language specific grammatical feature bundles (i.e. gender, number, declension class, etc.) to construct a language specific syntactic exo-skeletal frame. Following this perspective, a competent bilingual will have three sublists: a sublist of abstract morphemes from language A; a sublist of abstract morphemes from language B; and a third sublist of all known roots (Grimstad, Lohndal & Afarli, 2014). The development of these language specific sub-lists is part of the process of language acquisition. These abstract morphemes combine with roots to produce syntactic frames which are fully morphosyntactically specified before spellout. Notably, these feature bundles in the syntactic frame remain abstract and can be realized by any vocabulary item which matches the terminal nodes in the syntactic frame. As pointed out by Grimstad, Lohndal & Afarli (2014), this fact will prove crucial as intra-word codeswitching takes place purely at the level of vocabulary insertion.

Phonological exponents of lexical items are found in list 2; the Vocabulary. These phonological exponents, known as *Vocabulary Items*, are inserted into the derivation via a process called *Vocabulary Insertion*. As the syntactic frame has already been constructed and vocabulary items are morphosyntactically underspecified, phonological exponents are inserted late (i.e. post syntactically) into the derivation during spell-out²⁰. Crucially, as roots are not marked (or tagged) with language identity, this process allows Vocabulary Items associated with any language, dialect or register to be inserted into the derivation—this is known as free insertion²¹ (Grimstad, Lohndal & Afarli, 2014; Alexaidou et al., 2015).

Finally, in DM, the derivation is semantically interpreted by the conceptual-intentional system with the aid of list 3. The conceptual-intentional system directly interprets the syntactic features from list 1 but when semantic meaning cannot be derived from the morphosyntax alone, as is the case with lexical roots, the 3rd list, (i.e. the

²⁰ Evidence from Hopi supports the claim that root forms are inserted late (Bliss, 2004)

²¹ Haugen & Siddiqi (2013) argue for competition during the insertion of suppletive roots as opposed to free-insertion

Encyclopedia) provides the necessary semantic information. This list is the ‘list of special meanings’ (i.e. meaning not derived from the morphosyntax) for both individual root morphemes²² as well as larger phrasal idioms; presumably, this information is stored in LTM.

For example, consider the word *pen*. In List 1, the representation for *pen* will contain syntactic features such as [+noun], [+singular], etc. The second list contains the phonological form [pɛn]. The third list will hold a number of possible interpretations for the word *pen* and will select the appropriate one based on context. So, in the phrase *I have a blue pen* the word *pen* will be semantically interpreted as a writing instrument; however, if the word *pen* were to appear in a phrase *like the pen is mightier than the sword* where *pen* metonymically represents ‘writing’, then list 3 will produce an alternative interpretation best suited to the context. When discussing language mixing, specifically novel codeswitching or nonce borrowing, list 3 plays a vital role in interpreting root morphemes which may contain phonological exponents from multiple languages.

3.3.3. DM & Language mixing

Before extending this theory of DM to include language mixing it is prudent to clarify a few central assumptions. First, it is assumed that bilinguals have “functionally separate distributed lexicons” (Burkholder, 2017). This notion of functionally separate lexicons is akin to contemporary accounts of how multi-linguals acquire and represent language; notably the Multiple Grammars Model (See Roeper, 1999; Amaral & Roeper, 2014) which postulates sub-grammars to account for how multiple linguistic systems inhabit the mind of a single speaker. Central to the Multiple Grammars Model is the notion that new grammars are developed from re-configuring pre-existing grammatical principles in order to account for linguistic variation (e.g. language, dialect and register). The result is a number of sub-grammars which are active only in specific contexts. Similarly, in the model of DM presented here, ‘functionally separate’ implies that a bilingual speaker learns to match specific feature bundles to specific contexts (i.e.

²² See Marantz (1996) “‘Cat’ as a Phrasal Idiom” for a discussion on the ‘special meaning’ of morphological roots

sublists)—the same contexts which also select language (see section 2.4.5); this matching of feature bundles to contexts is the natural product of language acquisition. This results in specific abstract morphemes being associated with a specific language (or dialect or register); as roots do not contain language specific feature bundles they remain language neutral. Secondly, following the account of DM described above, lexical elements are not suppressed or inhibited based on language membership. As such, this paper will argue that abstract morphemes and roots (i.e. list 1) from either language A or language B (i.e. sublist A or sublist B) may be selected to build the syntactic frame, while Vocabulary Items (i.e. list 2) from both languages can be selected for insertion (Burkholder, 2017).

Additionally, following Grimstad, Lohndal & Afarli (2014), it will be assumed that in intra-word language mixing, one language, which provides the underlying syntactic frame, is considered the ‘host’ language; the other language which provides the phonological exponents may be referred to as the ‘donor’ language. These terms are used for ease of exposition and should not be confused with similar terminology found in the Matrix Language Model of codeswitching (Myers-Scotton, 1991, 2003). As intra-word codeswitching is concerned with word internal language mixing and words appear in the larger context of phrases and sentences, the term *host language* is meant to refer to the language of the larger sentential context. The greater syntactic frame is generated using feature bundles from the host language which also supplies the syntactic features for the element that is codeswitched.

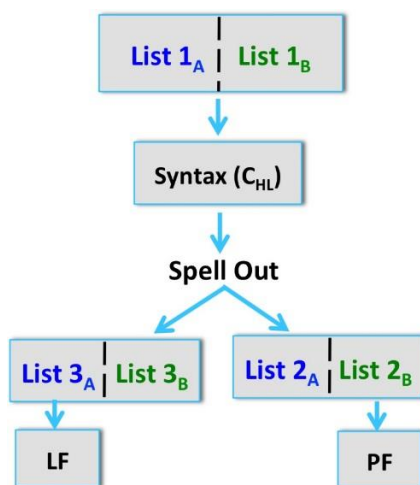
As pointed out by Grimstad, Lohndal & Afarli (2014), when a model of Distributed Morphology is applied to a bilingual/multilingual’s lexicon intra-word language mixing appears to be part of the natural fallout of language use; this is a sharp contrast to the lexicalist model proposed by MacSwan. In DM, while syntactic frames can only contain syntactic feature bundles from a single language, these frames are blind to the language identity of Vocabulary items. Phonological exponents from either and/or both languages may be inserted into the frame assuming they meet the syntactic demands of the exo-skeletal frame. This notion is characterized by Halle (1997), in the Subset Principle:

Subset Principle (Halle, 2000): The phonological exponent of a Vocabulary Item is inserted into a position if the item matches all or a subset of the features specified in that position. Insertion does not take place if the Vocabulary Item contains features not present in the morpheme. Where several Vocabulary Items meet the conditions of insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen.

As pointed out by Grimstad, Lohndal & Afarli (2014), this Subset Principle plays a vital role in constraining intra-word codeswitching. Crucially, during the production of a codeswitch, this allows for phonological exponents from any language to be inserted into the exoskeletal frame, regardless of the language identity of either the syntactic or phonological elements—assuming the vocabulary item which is inserted meets the demands of the active features in the syntactic frame. In a codeswitching context (i.e. codeswitching communicative mode), this allows functional vocabulary items from languages A and B to compete with each other. If the syntactic frame contains features [+X, +Y, +Z], any Vocabulary item matching these features or any subset of these features, (i.e. [+X, +Y], [+Y, +Z] or [+X, +Z]) may be inserted into the derivation. However, a Vocabulary Item may not be inserted if it contains additional features not present in the frame (e.g. [+X, +Y, +Z, +A] or [+X, +Y, +B]). This principle will prove crucial to accounting for when intra-language codeswitching may occur.

When considering bilingual speech, the illustration of DM introduced in Fig. 15 needs to be revised in order to account for multiple linguistic systems. This revised model is illustrated below in Fig. 16.

Figure 16: A DM Model of a Bilingual Grammar



(Burkholder, Mathieu, & Sabourin, 2017)

In this model of DM, elements from List 1, which are selected for the numeration by the C_{HL} , are drawn from a shared pool despite being largely associated with one language or the other (i.e. sub-grammar). The C_{HL} then constructs a language-specific syntactic frame from the exponents pulled from either list 1a or list 1b. During PF spell-out, Vocabulary Items from list 2a and list 2b compete against each other for insertion into the exoskeletal frame—phonological exponents from multiple languages may be inserted into the same syntactic frame. Finally, list 3 interprets the phonological exponents inserted into the syntactic frame and provides a logical form or meaning to the derivation.

3.3.4. Norwegian-English Language Mixing

To illustrate how this model of DM accommodates language mixing let us return to the Norwegian-English example introduced in chapter 1 and restated below in example 9. The analysis presented in this paper will closely follow work by Grimstad, Lohndal & Afarli (2014) and Alexaidou et al. (2015). As already mentioned, this data was collected from an American heritage community of Norwegian-English speakers and was drawn from the CANS (Corpus of American Norwegian Speech).

- 9) den **field**-a
 that field - DEF.F
 ‘that field’

Before such data can be fully examined there are a few aspects of Norwegian which must be first understood. To begin, Norwegian nouns are inflected with both number and definiteness and classified by gender. This paradigm is presented in table 3 below:

Table 3: Norwegian Nominal Affixes with Examples

Norwegian affix	Masc.	Fem.	Neuter
Indef + SG	en- en gutt (a boy)	ei- ei dor (a door)	et- et hus (a house)
Indef + Plural	-er gutter (boys)	-er dorer (doors)	[n/a] hus houses

Def + SG	-en gutten (the boy)	-a / -en* dora /doren* (the door)	-et huset (the house)
Definite + Plural	-ene guttene (the boys)	-ene dorene (the doors)	-ene / -a husene / husa (the houses)

(data from Wikipedia; corroborated with typecraft.org)

Of particular interest to this thesis is the [Def.SG.Fem] suffix *-a* which appears in our data. Like English, Norwegian nouns are introduced by determiners; unlike English, Norwegian determiners must agree with the *number*, *gender* and *definiteness* of the noun. This agreement between a nominal suffix and its determiner is known as *double definiteness* marking in that the DefP provides feature-values for underspecified features in the D. Double definiteness is a standard feature of Norwegian and is illustrated above in example (9), *den field-a*, via agreement between the definite determiner *den* and its corresponding suffix *-a*.

In their account of language mixing, Alexaidou et al. (2015) build upon an analysis by Julien (2009) who developed an abstract exo-skeletal frame for a Norwegian DP structure (see fig. 14 for the tree). In this frame, the root has no syntactic features of its own, rather gender is provided by the ‘n’ head; gender is part of the syntax not the root. Additionally, the operation Agree assigns unvalued features in ‘D’ values from their functional heads. This analysis also relies on the presence of a functional Def-head to account for double-definiteness marking. A Def-feature appears under the DefP which agrees with the unvalued Def-feature in the DP. This frame is seen below:

Norwegian DP Exoskeleton (Julien, 2009)

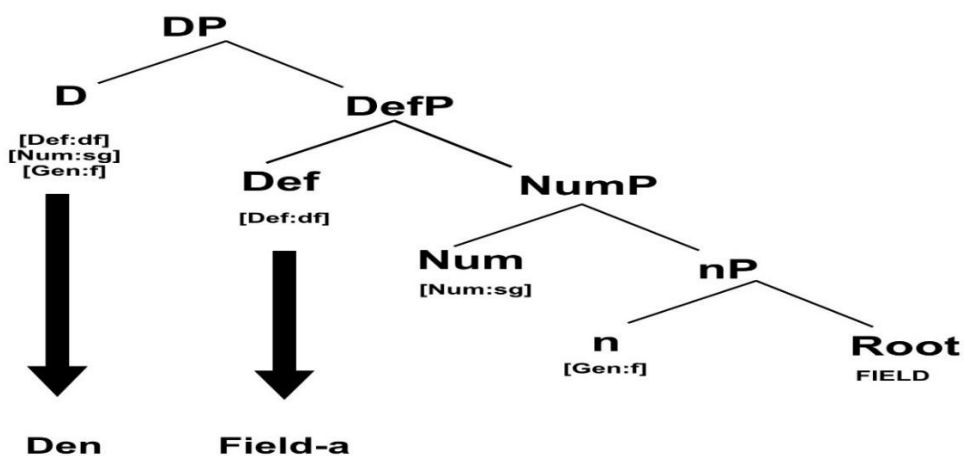
$$[DP D_{[DEF: U, NUM: U, GEN: U]} \dots [DefP Def_{[DEF: X]} [NumP Num_{[NUM: Y]} [nP n_{[GEN: Z]} \sqrt{ROOT}]]]]$$

In the language mixing analysis provided by Alexaidou et al. (2015), this exoskeletal frame functions as the underlying template into which lexical items are inserted.

Returning to our example, the syntactic structure for *den field-a* is illustrated below in fig 17. Note that the Vocabulary items *den field-a* have been inserted to make the derivation

easier to read; root vocabulary items are written in capitals. The determiner agrees with the [Def] feature of the DefP; this is double-definiteness marking.

Figure 17: Syntactic Structure for *den field-a*



(adapted from Alexaidou et.al., 2015)

In this example, the syntactic frame is generated using Norwegian features; the host language is Norwegian and the phrase *den field-a* is presumably part of a larger context where the Norwegian language dominates the conversation. As such, list 1 generates a syntactic frame using Norwegian features. This is evident by the presence of double-definiteness marking; the D head contains unvalued features which receive their values from the functional heads via Agree—in order to check the unvalued definiteness feature in the syntactic frame, D agrees with the Def head; likewise, D also agrees with the heads of NumP and nP in order to check the unvalued number and gender features.

English is the donor language. In List 2, English and Norwegian Vocabulary Items are in competition with each other for insertion into the syntactic frame. In this example, the English Vocabulary item, *field* is selected and is freely inserted into the Norwegian template. The root combines with the [Gen] feature which is introduced on the n head before merging with the Num head and the Def head respectively; the resulting feature bundle is morphophonologically realized as the codeswitch *fielda*. This occurs because the [Gen], [Num], and [Def] are realized during PF spellout as the Norwegian suffix *-a* while the English Vocabulary Item *field* is inserted into the root. It is noteworthy to point out that, while this approach allows for intra-word codeswitching to occur, it does not provide an answer as to why the English Vocabulary item—*field*—

was selected for insertion; in DM it's a matter of free insertion. This question will be addressed in chapter 4 where a production-oriented (i.e. performance-based) account of codeswitching will be developed.

3.4.0. Chapter Summary & Conclusion

This chapter began by introducing the phenomenon of codeswitching and discussing several types of language mixing. Of particular interest to this thesis is the phenomenon of intra-word codeswitching where elements from two languages merge to form a single word (i.e. $M_{lx} + M_{ly}$). Crucially in these examples of intra-word codeswitches, the 'switched element' (i.e. the donor element) maintains its original L2 mental representation—this is evident by the fact that donor elements in intra-word codeswitches preserve their original phonological form. Notably, this lack of phonological integration is regarded a defining feature of a codeswitch and allows researchers to distinguish between codeswitches and loanwords (i.e. borrowings).

The sociolinguistic implications of codeswitching were also briefly discussed. While there are a number of social and cultural factors which can motivate codeswitching, there is no universal or omnipresent impetus which can account for all instances of language mixing. Instead a number of researchers (Poplack, 1980; MacSwan, 2005; etc.) have suggested that bilinguals engage in a 'bilingual communicative mode' where the act of codeswitching itself carries meaning—functioning as a 'verbal wink' of sorts.

Notably, a number of researchers investigating codeswitching have attempted to identify syntactic boundaries which constrain codeswitching. However, as ultimately seen in Table 2, there is no consensus on universal boundaries constraining codeswitching. This lack of agreement amongst language mixing researchers led MacSwan (2005) to claim that 'nothing constrains codeswitching apart from the grammars involved—he formalizes this claim as by stating that codeswitching is the union for *grammar X* and *grammar Y* (i.e. codeswitching = $\{G_x \cup G_y\}$).

However, MacSwan's lexicalist approach to codeswitching contains a prohibition on intra-word codeswitching known as the PF Disjunction Theorem. This theorem argues that as languages have language-specific hierarchically-ordered phonological rules, and as these rules vary cross-linguistically, it is impossible for a single lexical item

to contain phonological elements from more than one language as there is no way to merge two sets of ordered rules during PF spellout; any attempt to do so will cause the derivation to crash—this results in MacSwan’s prohibition on intra-word codeswitching.

As such, this thesis turned to a DM model presented by Grimstad, Lohndal and Aferli (2014) to provide a structural account intra-word codeswitching. Crucially, distributed morphology provides a molecular view of a word similar to what is found in the MOGUL framework (see Chapter 2); a word is a composite of elements formed from merging features from three different lists (i.e. terminal nodes, vocabulary, encyclopedia). Crucially, when discussing language mixing, distributed morphology allows for disparate elements from different languages to come together over the course of single derivation, thus accounting for intra-word codeswitching.

Significantly, Distributed Morphology offers a competence based, representational account of language which focuses on the formation of grammatically sound derivations—DM is interested in I-language. This fact will prove crucial in chapter 4 when this model of DM is placed inside a performance-based model of cognition (i.e. MOGUL); as noted in chapter 1, a central goal of this thesis is to exploit a DM model of I-language to inform the production of intra-word codeswitching in the MOGUL framework.

Chapter 4: A Unified Framework

4.0.0. MOGUL and Intra-word Language Mixing:

A central goal of this chapter is to demonstrate that language mixing, specifically intra-word codeswitching, can be analyzed as natural consequence of the MOGUL architecture. This follows Sharwood-Smith and Truscott (2016) who note that, somewhat paradoxically, “a theory of codeswitching should, ideally, not be a theory of codeswitching”. Instead, any cognitive account of codeswitching should focus on a more general theory of representation and processing which can then be applied to language mixing phenomena. Ultimately, this approach allows for the language mixing phenomenon to be accounted for using general cognitive and linguistic theories—no extra machinery is necessary. To this end, a number of cognitive and linguistic apparatuses have been introduced. Prudence dictates a brief review of the key elements introduced in chapters 2 and 3 before they are brought together to account for intra-word codeswitching within the MOGUL framework.

4.1.0. The Cognitive Framework: Review

In chapter 2, I presented the MOGUL framework (SS&T, 2014). This framework is concerned with modeling linguistic representation and processing within a modular architecture in real time. MOGUL adopts a Jackendovian-style modular architecture. Of central interest to this thesis are the PS (Phonological Structure) and the SS (Syntactic Structure) which constitute an encapsulated language core; these components construct phonological and syntactic representations from module specific primitive features or feature bundles. While these primitives are language-neutral they may be bundled together into context-specific (i.e. language-specific) combinations of features. The representations from the language core interface with the CS (Conceptual Structure) which is largely responsible for semantic and pragmatic representations. The result of this interfacing is the formation of representational chains, such as [PS + SS + CS], which are the MOGUL counterpart to lexical items. Crucial to this account of intra-word codeswitching, these representational chains may be formed from module-specific representations which are associated with a combination of different contexts (i.e. languages).

In MOGUL, a set of variety-specific linguistic representations (i.e. a language-variety, dialect or register) receives an activation boost through the process of conceptual triggering. This process occurs when a ‘General Language Representation’ (GLR), which is part of a complex CS representation, is activated by cognitive context. As the ‘general language representation’ is co-indexed to all feature bundles associated with it, all co-indexed representations receive a corresponding increase in their activation levels. It is worth noting that non-conceptually triggered representations (i.e. non-dominant languages) are still in regular competition for selection; however, conceptually triggered representations receive an additional boost in activation which allows them to win the competition. This activation boost only affects feature bundles associated with a specific language variety; even though representations associated with different languages are stored in the same structure (e.g. all SS representations are in the SS) context selectively determines which representations are active. The result yields the appearance of separate lexicons which are activated via context despite all representations being stored in the same place.

Cognitive context is also accounted for via modular representation. The CS as well as both the POpS (Perceptual Output Structures) and AffS (Affective Structures) contribute to building a cognitive context. POpS interpret sensory information to form an internal representation from an external context (i.e. the real-world environment), while AffS associate value representations with both contextual and linguistic representations—a representation with a value representation will have an increased resting level of activation equal to the strength of the co-indexed value representation (i.e. a strong value representation will result in a relatively large increase in activation levels for co-index representations). While the CS, POpS and AffS work together to establish context, it must be noted that the whole story of cognitive context is much more complex than what has been present in this thesis and needs to take into account associated episodic and procedural memories as well as personal experience; all of which can be postulated to function in a modular fashion consistent with the MOGUL architecture but whose nature is outside the purview of this thesis.

In addition to providing semantic and contextual information the CS is also responsible for establishing goal representations. These goal representations interact with

cognitive context—most notably AffS, which marks the most salient goals with high value representations. According to the MOGUL model, goals with high value representations motivate and direct action. As goal representations are part of general cognition they include general motivators like ‘quell hunger’, ‘sleep’ or ‘find a bathroom’; more relevant to this story of language-mixing are social-communicative goals, which include representations that potentially motivate intra-word codeswitching (e.g. communicate a message, foster solidarity, humour, etc.). When goal representations promoting codeswitching are active, a speaker engages a bilingual mode of communication where it is the act of codeswitching rather than the language itself which carries its own meaning—akin to a gesture.

It is worth emphasizing that a speaker does not enter ‘bilingual mode’ because codeswitching is the goal; rather ‘bilingual mode’ is a reflection of cognitive context while codeswitching is a task the mind/brain completes in order to satisfy the goal. There may be several different tasks which can satisfy any given goal. In his work on consciousness and language learning, Truscott notes that “A goal is a component of SELF that represents a desired state. Action plans compatible with the achievement of this state are active when the goal is activated” (2014, p. 89). So, a goal like, ‘quell hunger’ can be satisfied by eating food or smoking a cigarette (i.e. a chemical appetite suppressant); likewise, a social goal like ‘display special relationship’ (e.g. when an individual wishes to highlight a special relationship) might be satisfied by codeswitching, telling an inside joke or an esoteric gesture. When an individual produces an intra-word codeswitch, they do so because ‘bilingual mode’ is active and codeswitching is the best way to satisfy their goal.

Finally, a cornerstone of the MOGUL framework is APT (Acquisition by Processing Theory). APT provides an account of growth and learning within the MOGUL architecture. The central tenet of this theory claims that learning is the lingering effect or byproduct of processing. All representations have a resting level of activation; when a representation is stimulated its activation level will spike and return to a slightly higher resting level. Frequently activated representations have a high resting level of activation which, in turn, gives these representations a competitive advantage when vying for selection. When a speaker learns a new word, or creates a neologism, a

novel [PS+SS+CS] chain is created and stored in long-term memory. The newly formed representational chain will have a low resting level of activation. Each time the speaker activates the new chain, it increases its resting level of activation in addition to developing new contextual associations. This accounts for how new lexical items are formed and how they come to be associated with specific language varieties (i.e. GLRs).

4.1.1. The Linguistic Framework: Review

Linguistic frameworks for codeswitching were the focus of chapter 3. Socio-linguistic motivations for codeswitching were also discussed. While researchers have offered a number of motivating factors (e.g. filling lexical gaps), Poplack (1980) claims that language mixing amongst bilingual communities is the result of a ‘bilingual speech mode’. In such circumstances, the *act* of codeswitching carries meaning, functioning like a gesture, codeswitching may occur to foster solidarity, draw attention to specific information, initiate humor, or serve numerous other social functions.

Two Minimalist syntactic accounts of language mixing were discussed. A lexicalist approach (MacSwan, 2000, 2005) and a DM approach (Grimstad, Lohndal & Afarli, 2014) were both briefly reviewed. These frameworks offered null theories of codeswitching—accounts of language mixing which do not postulate any extra or special machinery to accommodate codeswitching. Significantly, by adopting a null theory, codeswitching can be explained via a general account of language—an account of codeswitching that follows the same linguistic procedures as monolingual speech, but involves two language varieties.

MacSwan’s lexicalist approach to language mixing was reviewed in some detail. Crucial to this account of language mixing is MacSwan’s assertion that nothing constrains codeswitching apart from the grammars involved. This means that there are no universal syntactic constraints specific to codeswitching; language mixing may occur at any point in a structure (e.g. sentence, phrase, constituent) as long as it is consistent with the specific grammars involved. However, within MacSwan’s lexicalist approach to language mixing, the PF Disjunction Theorem²³ places a prohibition on intra-word

²³ Note the PF Disjunction Theorem is restricted to the word level and does not apply to sentences. As phonological rules are applied at the sentence level as well as the word level, it is unclear why phonological systems can be mixed at the sentence level but not the word level in MacSwan’s model.

codeswitching due to C_{HL} 's inability to merge ordered lists of phonological rules.

However, as the PF Disjunction Theorem is born from the implications of a Lexicalist approach (i.e. the Lexical Integrity Principle: Bresnan & Mchombo, 1995) this prohibition does not hold true for a DM account of language mixing.

As such, this thesis has adopted a model of DM codeswitching following work by Grimstad, Lohndal, and Afarli (2014) and Alexaidou et al. (2015). Under this approach, a fully specified morpho-syntactic frame is constructed using syntactic terminal nodes (i.e. list 1). Vocabulary items—containing phonological information (i.e. list 2)—which match the feature bundles in the frame are inserted, post-syntactically, into this frame in a process known as late insertion. As the structure is being realized by Vocabulary items, the derivation is interpreted via the Encyclopedia (i.e. list 3) during LF spellout.

When considering language mixing, DM has several advantages over the traditional lexicalist approach. To begin, DM postulates that words have internal syntactic structure and are constructed by merging lexical roots with abstract morphemes—a hierarchical structure all the way down driven by a single engine. Lexical roots are category-neutral, and are not associated to any specific language (Grimstad, Lohndal & Afarli, 2014); abstract functional morphemes are inserted into syntactic feature bundles which are in language-specific configurations (Grimstad, Lohndal & Afarli, 2014). As the root is syntactically separate from affixal morphology there is opportunity for language mixing to occur during Vocabulary insertion. This paper will argue that codeswitching is the product of a fully-specified morpho-syntactic frame which has Vocabulary items from two languages inserted into it.

The notion of late insertion is crucial to account for intra-word language mixing. In DM, this refers to the insertion of Vocabulary items (i.e. phonological form) into a fully-specified morpho-syntactic frame. Following Grimstad, Lohndal & Afarli (2014), I assume that the host language generates the syntactic frame using language-neutral roots and language-specific abstract morphemes. If the syntactic template is blind to the language identity of items in list 2, Vocabulary items from either language may be inserted into the frame, assuming they satisfy the Subset Principle (see section 3.3.3). Intra-word codeswitching is the result of a root and its abstract morpheme being satisfied by Vocabulary items from two different languages.

Additionally, the three lists adopted in DM have a remarkable similarity to the molecular view of a word postulated in MOGUL. Like MOGUL, DM divides the content of a lexical item amongst multiple domains (i.e. lists). While DM is somewhat vague as to what these lists are and how they represent and store information, MOGUL gives them shape in the form of modules. In DM, list 1—together with the syntax—functions similarly to the SS which establishes syntactic representations; the operations of list 2 during PF spellout correspond to the PS which interfaces phonological form with a syntactic representation. In DM, the conceptual/intentional system interprets the derivation while list 3 provides access to special meanings which are not interpretable from the morpho-syntax. In MOGUL the CS is the locus of semantic meaning; ‘special meanings’ (e.g. idioms, metaphors) are complex associations (i.e. a series of co-indexed representations) which emanate from the CS and are stored in LTM; the conceptual/intentional system found in DM is largely paralleled in MOGUL by the interactions between the CS, AffS, goal representations and cognitive context. As such, a lexical item in DM (e.g. List 1 + List 2 + List 3) is akin to a lexical item in MOGUL (e.g. SS + PS + CS). This thesis will argue that this inherent compatibility regarding the molecular structure of a word allows a DM account of language mixing to function within the MOGUL framework.

4.2.0 Situating DM in a MOGUL Framework: Bringing it all together

This account of codeswitching begins with cognitive context. An internalized mental model is constructed by the speaker to reflect the external environment they are experiencing. In addition to physical location, internal factors also influence cognitive context, including factors such as the speaker’s self-image; their knowledge of and relationship to the interlocutor; personal motivations (i.e. goals); and the value representations they attribute to various aspects of the context. As these factors are mentally updated in real time, cognitive context is fluid which responds in real time and can change its shape from moment-to-moment. Within the MOGUL architecture, these factors manifest in the form of representations and their tangential associations. A number of modules contribute to constructing a cognitive context including the CS, POPS and AffS—experience (e.g. episodic memory & procedural memory)) also plays a role (SS&T, 2014) but falls outside the purview of this thesis. Active representations, and

associated representations that receive tangential activation in these modules, constitute context, which results in conceptual triggering of a specific language in a specific situation.

The language which is conceptually triggered is said to be the dominant language; all representations that are associated (i.e. co-indexed) with the triggered language will receive an additional activation boost; the additional activation boost means representations co-indexed to the conceptually triggered language will usually dominate any competition. When a speaker experiences a cognitive context which conceptually triggers the dominant language, the result is the construction of a syntactic frame when the speaker plans an utterance.

In DM, once the syntactic template has been constructed, Vocabulary items compete for insertion. In order for intra-word codeswitching to occur, a speaker must engage a ‘bilingual mode’ of communication. In MOGUL, DM vocabulary items correspond to PS representations (and their co-indices to SS representations). Under most circumstances, the PS representation inserted into the syntactic frame is the PS representation which best matches representations in both the SS and the CS, and which has been conceptually triggered. However, it is argued here that, in language-mixing situations, cognitive context is oscillating between two dominant language contexts. This notion of oscillation is a crucial one; only one language can be dominant (i.e. conceptually triggered) at a given time; however, context can change in real time. The oscillation between two languages occurs as the result of real-time alteration to cognitive context—these changes may be external (i.e. changes to the physical environment) or internal (i.e. re-evaluating goals or value representations).

The metaphor of contextual oscillation represents a rapid alternation of dominant linguistic contexts. As language is constructed in real time and, neurologically, activation levels are measured over a period of milliseconds²⁴, this fluctuation between dominant contexts can, in principle, happen a number of times over the course of constructing a single word. Following MOGUL assumptions (SS&T, 2004, 2014) regarding representations & processing (see section 2.3.0 on APT for details),

²⁴ Neural activations are measured in amplitude (i.e. microvolts) over time (i.e. milliseconds).

representations associated with the dominant language (e.g. language A) will receive an activation boost; when the dominant context switches to language B, high activation levels for language A will linger a moment before they decay; when the context switches back to language A these decaying representations increase their activation levels again accordingly. If the context switches back and forth between language A and B in rapid fashion, representations may not be given the opportunity to decay before they are boosted again. The result yields the appearance of two languages being simultaneously conceptually triggered; in truth, a constant array of minor changes made to a single cognitive context cause it to rapidly alternate between triggering two different languages, boosting them both. Crucial to this account of codeswitching, it is the conceptual triggering of two general language representations (i.e. GLRs) during the processing of a single derivation which allows PS representations from two different languages to compete against each other. It is this state of oscillating dominant contexts, where representations from two languages both receive an activation boost from conceptual triggering, that I claim constitutes the basis of a ‘bilingual mode’ of communication; this account functions as an operationalization of communicative mode in MOGUL, which is the natural by-product of how context operates in the MOGUL framework.

When a speaker engages in ‘bilingual mode’, representations from two languages are able compete against each other. A DM framework would allow for Vocabulary items from both languages to compete to be inserted into the syntactic frame. Any Vocabulary item which fulfils the subset principle may be selected (i.e. Vocabulary items must meet the set or subset of features present in the syntax). I propose that within the MOGUL framework the subset principle is accounted for as a natural by-product of co-indexation. The subset principle is accounted for because the activation level of any PS representation will be determined by the activation levels of any co-indexed representations in other modules—in MOGUL co-indexation between module-specific representations results in a mutual increase in activation for said representations. If an SS feature bundle increases in activation, PS representations with co-indexed features will also increase in activation; the PS representation with the most features co-indexed to the

active SS representation will be the most active in the PS module²⁵ and will be inserted into the derivation; this is competition in MOGUL. As such, feature co-indexation, which increases the activation level of all co-indexed PS representations, satisfies the subset principle; features in the PS do not match or satisfy SS features (i.e. there are no syntax features in the PS), but rather co-indexation ensures that syntactic features activate the appropriate PS representations—co-indexation is developed over time via experience and exposure to language (see chapter 2 section 2.3.0 on APT for details). This means that any PS representation which is co-indexed to an SS representation will, by the very nature of co-indexation, meet the syntactic demands of the SS feature bundle.

During the processing of a derivation in MOGUL, lexical items are formed from the most active representations in each module which are co-indexed together. When considering an instance of codeswitching, each feature bundle in the SS is potentially co-indexed to a pair of PS representations; one from the host language and one from the donor. In a language mixing situation, a speaker has two general language representations active in their CS (i.e. conceptually triggering two languages). Representations co-indexed to either general language representation (i.e. GLRs) will receive an additional activation boost. Effectively, this means that PS representations associated with two different languages are competing on an equal playing field; from here competition may proceed in the standard MOGUL fashion.

However, the question remains: when elements from the host and donor language are in competition, what allows the donor language to win? In standard bilingual environmental circumstances (i.e. oscillating contexts), when PS representations from the host and donor language are in competition, one would predict that the host representation should always win the competition as previously attested (i.e. experienced) representational chains will have higher resting levels of activation than newly formed chains; even on a level playing field the host language should still win. I argue that this phenomenon can be accounted for via the construct of goal representations. Communicative goals contribute to cognitive context but also increase the activation level of representations which satisfy their demands. When a speaker engages a

²⁵ Note that, the PS representation cannot have features which are not co-indexed to the SS; this would violate the subset principle

‘bilingual mode’, communicative goals conducive to codeswitching will be active in the CS. In a language mixing situation, this translates to an increase in the activation level of the donor language representations in order to satisfy the goal. As such, goal representations play a central role in permitting a speaker to mix languages.

In a standard DM derivation, once Vocabulary items have been inserted into the syntactic frame, the derivation receives semantic interpretation from conceptual/intentional system during LF spellout; if a derivation carries a ‘special meaning’ (i.e. a root or idiom) then List 3 is consulted. But how does the encyclopedia function in a MOGUL framework? In MOGUL, semantic content is co-indexed to syntactic content; an increase in the activation level of an SS feature bundle will subsequently increase the activation level all associated CS representations in real-time. For example, the syntactic features which represents the noun *dog* will also be employed to represent functionally similar nouns like *cat* or *mouse* (Harley & Noyer, 2000). In MOGUL, when the feature bundle co-indexed to *dog* becomes active in the SS, all co-indexed CS representations (e.g. *cat*, *mouse*) will receive an activation boost. However, a good deal of semantic information is represented outside the CS. The selection of a specific lexical item—*dog* or *cat*—depends on the active contextual elements interfaced to CS. While *dog* and *cat* share SS features and basic CS representations (e.g. theta roles, number, general language representation, etc.), complex CS representations have additional co-indexations to POpS and other extra-linguistic modules. Any visual information that contributes to the meaning of a lexical item (i.e. colour, shape, size, etc.) is stored in the VS and is interfaced to the CS. Likewise, all sensory information (AS, OS, GS, TS/KS) to the extent that it contributes to meaning, is interfaced from POpS to a complex CS representation. In this way, sensory information stored outside the CS participates in the development of complex meanings which have co-indexed syntactic and phonological form. The formation of these complex CS representations allows for the semantic differentiation of lexical items. This process of building complex CS representations (i.e. constructing a lexical meaning) by co-indexing the CS representations to extra-linguistic modules follows the same process regardless of language mode (e.g. monolingual mode, bilingual mode, etc.). Crucial to this thesis, it is this view of how semantic meaning is developed and represented in the mind/brain which

accounts for, at least in part, how the DM Encyclopedia can be represented within the MOGUL architecture.

4.2.1 Intra-word Codeswitching in MOGUL: An Empirical Example

To illustrate the account of intra-word codeswitching elaborated on above, I will return to the English-Norwegian language mixing data introduced in chapter 1 and again in chapter 3. These examples were taken from the CANS corpus of heritage Norwegian speakers (Grimstad, Lohndal & Afarli, 2014), a community of English-Norwegian bilinguals living in the north-central United States. Example (10) is repeated below:

- 10) den **field**-a
 that field-DEF.F
 ‘that field’

Example (10) is a phrasal constituent which is presumably part of a larger phrase, sentence or conversation. Regardless of the exact circumstances, the dominant language in this context is assumed to be Norwegian which has been conceptually triggered via a specific context. As language users must always have a dominant language context, and as the syntactic frame is established via the dominant context, the individual who produced example (10) must have a Norwegian-dominated context because the syntactic frame was built from Norwegian features—this is evident from the Norwegian double definiteness marking²⁶. This cognitive context will reflect a number of internal (e.g. self-image, inter-personal relationships, etc.) and external influences (e.g. location, identity of interlocutor, etc.) which the speaker has co-indexed with the Norwegian general language representation in the CS. This means that all modular representations which are interfaced to the Norwegian general language representation will receive an activation boost that subsequently causes these representations to dominate any competition. As such when a speaker chooses to initiate language, a syntactic frame will be constructed from feature bundles associated with Norwegian. This analysis is taken from Alexaidou et al. (2015) who argue that in example (10), the speaker constructs a Norwegian DP frame; featured below (see chapter 3 for details):

²⁶ This analysis is based on work by Grimstad, Lohndal & Afarli (2014) and Alexaidou et al. (2015).

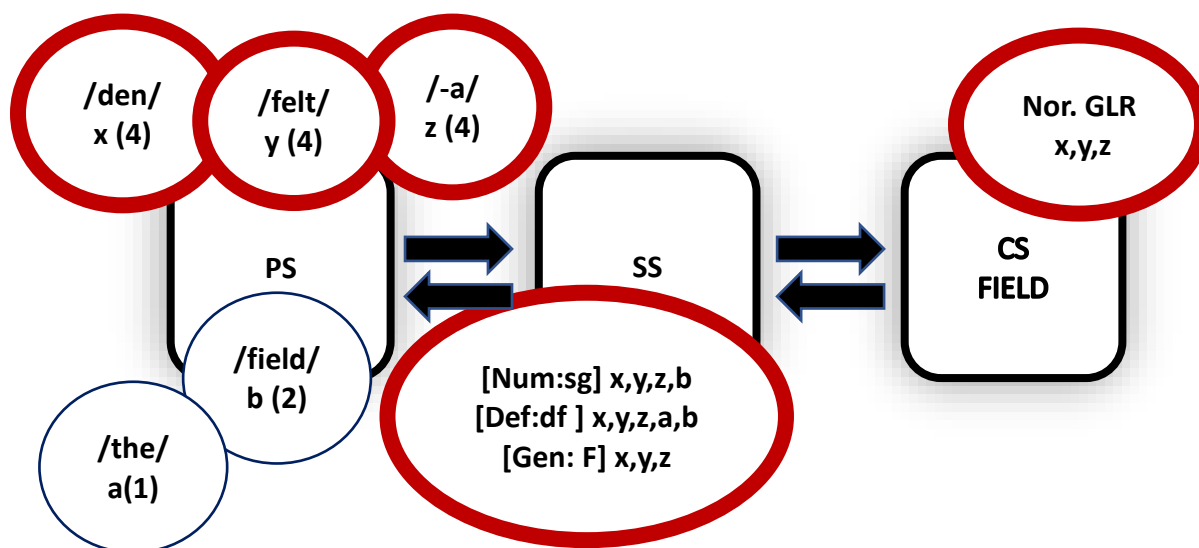
Norwegian DP Exoskeleton (Julien, 2009)

[DP D_[DEF: U, NUM: U, GEN: U] ... [DefP Def_[DEF: X] [NumP Num_[NUM: Y] [nP n_[GEN: Z] VROOT]]]

As in MOGUL, derivations take place in real-time, real-time alterations to a cognitive context have consequences; metaphorically, context is fluid. In monolingual circumstances, where Norwegian remains the dominant language, all representations in any module which are co-indexed to the Norwegian general language representation will have an increased activation level and dominate any competition for insertion. Feature co-indexation in a Norwegian-dominated context is pictured below in fig. 18. In this figure, lower case letters represent co-indexations; each co-indexation a representation has increases that representation's activation level—the bracketed number in the PS module²⁷ is the activation level which is based on the sum total of co-indexed forms (i.e. the number of tokens per type). For ease of exposition every co-indexation is weighted the same (i.e. boosts activation by 1). In this simplified illustration, Norwegian PS representations are co-indexed to three SS features and the general language representation (i.e. GLR) in the CS which results in an activation level of (4) for Norwegian PS feature bundles—note that even without an active English GLR, English features co-indexed to SS features still compete for selection but their relatively lower levels of activation mean they are not competitive. The output of fig. 18 is a monolingual Norwegian production – *den felta*.

²⁷ Activation levels are displayed in the PS to highlight changes that account for intra-word codeswitching; each co-indexation (i.e. x, a) will have the same level of activation in each module.

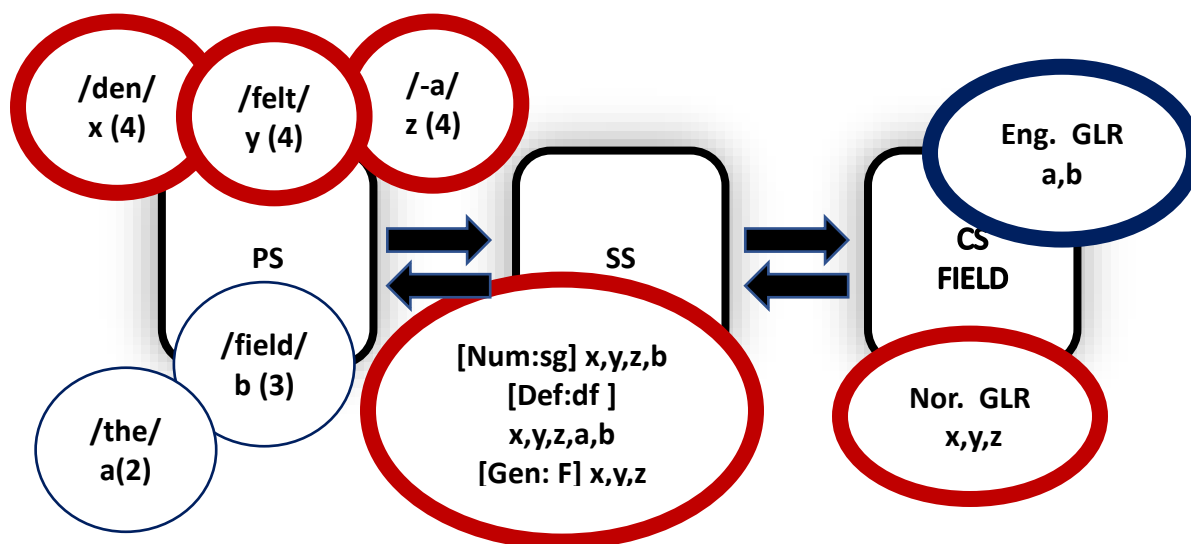
Figure 18: Feature Co-indexation



Legend : Bold = most active representation; red = Norwegian; blue = English;
 round square = module; circle = feature bundle

In language mixing circumstances, as in example (10) (i.e. *den field-a*), a speaker engages a bilingual communicative mode. Constant alterations to cognitive context—dependent on internal (e.g. goal or value representations) and/or external factors (e.g. changes to the physical environment)—cause the dominant language to oscillate between English and Norwegian. In MOGUL, this is realized as the alternation between peak activation levels of the English and Norwegian general language representations (i.e. GLRs) in a complex CS representation. As the context alternates between dominant languages, activation levels of representations associated with both English and Norwegian representations adjust accordingly. This constant adjustment of activation levels allows for English PS representations to be more competitive when fighting for insertion against their Norwegian counterparts. This oscillation is pictured below in fig. 19 which is based on fig. 18 but with the inclusion of an English GLR. When the English GLR becomes active the English PS representation receive an additional activation boost. However, as the Norwegian PS features are co-indexed to a greater number of SS features they will maintain a higher activation level; as such, the output of fig. 19 will be the monolingual Norwegian production—*den felta*.

Figure 19: Feature Co-indexation with Oscillating Contexts



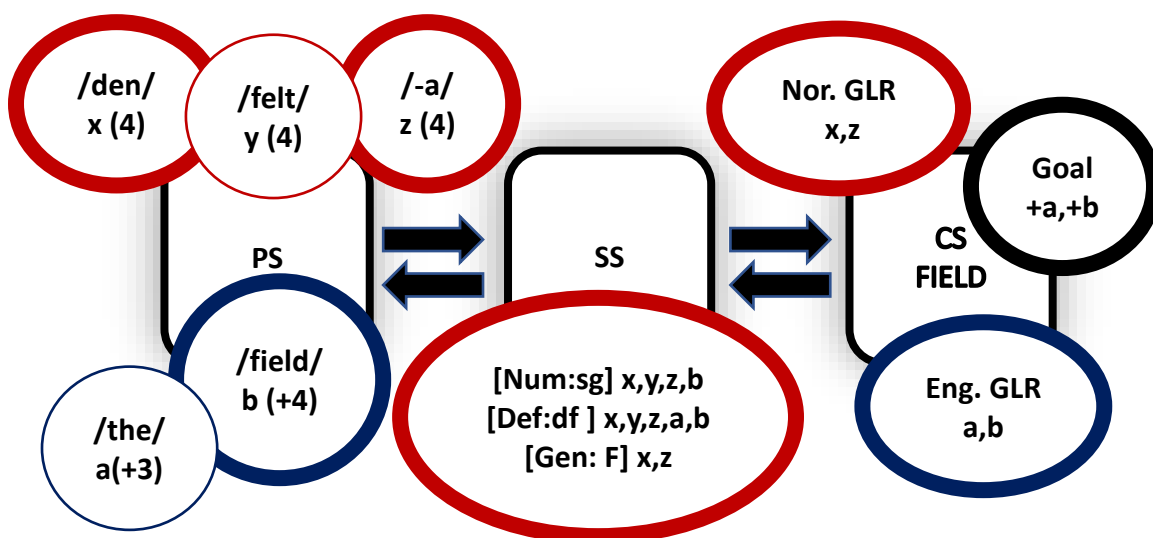
Legend : Bold = most active representation; red = Norwegian; blue = English;
round square = module; circle = feature bundle

In example 10, the initial Norwegian context, which established the syntactic frame, will increase the activation level of Norwegian PS feature bundles for *[den-felt-a]*. When the dominant language oscillates, English PS features, representing the lexical item *field*, increase in activation, and compete to be the most active item in the PS item representing the root. However, in order for the English PS item to become the most active it requires an additional activation boost which appears in the form of a goal representation. That is to say, while ‘bilingual mode’ allows for representations from two languages to be relatively competitive, it is co-indexation to a salient goal representation (i.e. the most active goal) which provide linguistic representations with an additional activation boost and cause them to be the most active elements in their respective modules.

Goals are part of a complex CS representation; any representation which is co-indexed to a goal receives an additional activation boost. Salient goals are co-indexed to high value representations in the AffS; the higher the value representation in the AffS the greater the activation boost. In example (10), a high priority goal representation is co-indexed to English PS items; the goal representation is co-indexed to a high value representation which indicates a salient goal that is indicative of a large activation boost—as goal representations drive thought and action, this thesis proposes that

representations co-indexed to a salient goal receive a ‘super boost’. This ‘super boost’ provides a greater increase to activations levels than regular co-indexed representations; in fig. 20 ‘super boosted’ representations are marked with (+) to indicate an additional boost, (i.e. $+4 > 4$). As the English PS representation is co-indexed to a goal it receives an activation boost of (+1) which raises its activation level to (+4) and makes it the most active PS feature bundle representing the root FIELD. The most active features in each module are then chained together and result in the English-Norwegian codeswitch, *den field-a* as seen below in fig. 20.

Figure 20: Feature Co-indexation with Oscillating Context and a Goal



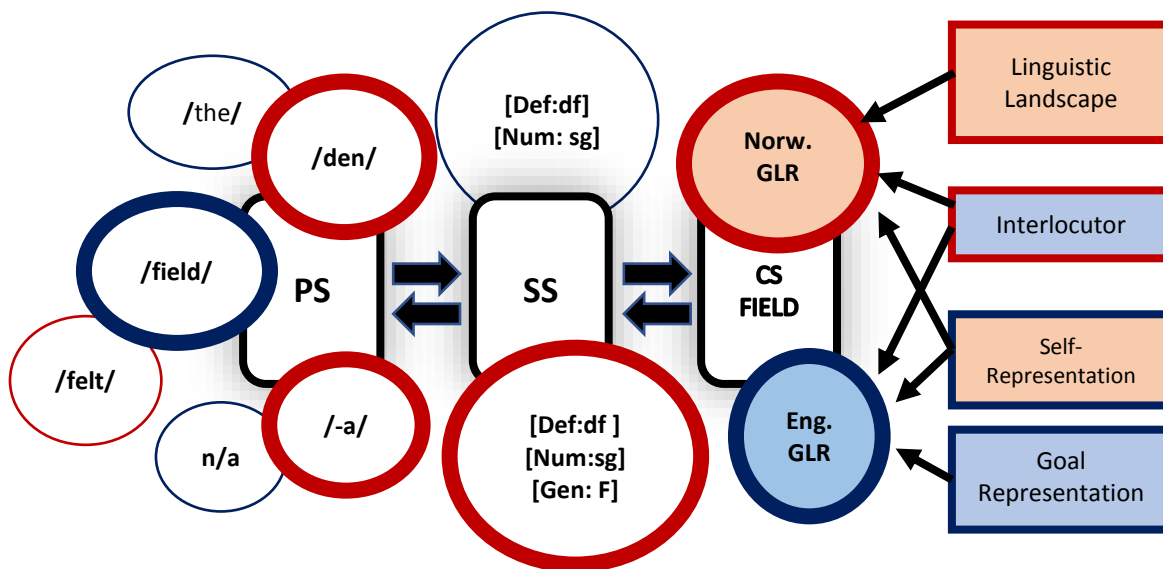
Legend : Bold = most active representation; red = Norwegian; blue = English; round square = module; circle = feature bundle

Note that the English PS representations for determiners are not co-indexed to the gender feature present in the syntactic frame. As Norwegian PS representations for determiners and roots are co-indexed to gender in the SS, they can maintain a higher level of activation even as the dominant language oscillates. The result is that representations for an English root and Norwegian determiner are the most active in the PS and are interfaced to other modules accordingly. This allows for an English lexical root (i.e. field) to be introduced by a Norwegian determiner and carry a Norwegian suffix.

An oscillating context combined with MOGUL’s incremental and dynamic processing perspective means modular representations from both languages are in competition during every stage of the derivation. Crucially, this allows for established

representational chains to be updated and altered in novel ways; thus, accommodating codeswitching. To illustrate, I have (partially) modeled this process below in fig. 21. Starting on the right-most side of fig. 21, contextual factors are associated with language; some factors like the identity of the interlocutor or a self-representation may be co-indexed to multiple languages—this is represented by multi-coloured contexts.

Figure 21: Derivation for *Den Fielda*



**Legend: bold = most active; red = Host/Norwegian; blue = Donor/English;
 round square = module; circle = feature bundle; rectangles = contexts**

This mixed language context causes the conceptual triggering of two languages in a near simultaneous (or oscillating) fashion which in turn allows representations from English and Norwegian to compete against each other. In fig. 21, the bolded circles represent feature bundles which are the most active in their respective modules; the most active feature bundle in each module is interfaced to form a (PS + SS + CS) representational chain. The result is intra-word codeswitching; the formation of representational chains which contain feature bundles associated to two different languages. These representational chains are illustrated below:

[(PS/PS/PS) + SS + CS/CS] or [(den/field/a) + (Def:df; Num:Sg; Gen:F) + (CS/CS)]

This process allows for the creation of novel combinations of mixed language representational chains. As MOGUL derivations do not crash, but are formed from the

most active content in each module, the PF Disjunction Theorem, as present by MacSwan (2005), is not applicable.

This approach to intra-word codeswitching demonstrates that this phenomenon is a natural consequence of the MOGUL architecture—intra-word codeswitching is the by-product of the real-time interaction between an oscillating context and the representations associated with said context. This processing approach is largely consistent with a Distributed Morphology model of syntactic derivation—notably, both MOGUL and DM share a molecular view of a word where the lexical content is distributed over a series of modules which are interfaced together.

It should be noted that MOGUL and DM approach language research from fundamentally distinct perspectives. DM offers a competence based, representational account of language which focuses on the formation of grammatically sound derivations from an atemporal perspective. However, MOGUL is primarily concerned with real-time performance and language use; the focus is on the production of grammatically acceptable derivations which transmit the desired meaning. By situating a competence-based model of grammatical structure (i.e. DM) inside a performance-based model of cognition (i.e. MOGUL) this thesis is able to account for how I-language informs the production codeswitching.

4.3.0 Discussion: MOGUL Predictions on Intra-word Codeswitching

Given the nature of the MOGUL architecture and MacSwan's assertion regarding the behaviour of codeswitching (i.e. nothing constrains it apart from the grammars involved), very few universal predictions can be made regarding intra-word codeswitching. As intra-word codeswitching is the product of cognitive context—bilingual mode—the number of syntactic predictions that can be made are limited.

Nonetheless, MOGUL does make a few assertions regarding the formation of intra-word codeswitches. As elaborated above, the very nature of cognitive context means that at any point in time an individual has a dominant language context. When considering intra-word codeswitching the language which is dominant in the greater sentential context will always be taken as the host language which supplies the syntactic frame—as contexts oscillate, PS representations from the donor language increase in activation and compete for insertion. This produces representational chains like, [PS_D +

$SS_H + CS_D/CS_H$ (D = donor, H = host), which account for intra-word codeswitching. Conversely, it is impossible for the MOGUL framework to produce an intra-word codeswitch where the SS representation is the donor element, as in $[PS_H + SS_D + CS_D/CS_H]$; this is because the host language constructs the syntactic frame for the entire phrase, constituent or sentence.

For example, consider the English phrase *the field* as a hypothetical English/Norwegian codeswitch where English is the host language and Norwegian is the donor. If the $[PS_H + SS_D + CS_D/CS_H]$ chain for *the field* contained a Norwegian gender feature in the SS, the gender feature is not co-indexed to any English PS representation; as such Norwegian gender is not marked. This type of codeswitching is impossible in the MOGUL framework the English PS features will not achieve sufficient levels of activation. If the gender feature of the Norwegian determiner is activated in the SS then the corresponding Norwegian PS representation (e.g. *den*) will have the greatest number of co-indexed features and will be selected over the English determiner (e.g. *the*). As such, it is not possible to produce a codeswitch like *the field* where the Norwegian SS features come from the donor language and the English PS representations are from the host language.

Alternatively, consider the Norwegian phrase *den felt* as a hypothetical Norwegian/English codeswitch where Norwegian is the host language and English is the donor. This type of codeswitching is ruled out by restrictions on PS representations (i.e. Vocabulary Items) requiring co-indexation to a subset of SS features (i.e. a violation of subset principle). If the $[PS_H + SS_D + CS_D/CS_H]$ chain for *den felt* contained only English representations in the SS there would be no gender feature co-indexed to the PS. This is problematic as Norwegian lexical roots require a gender feature to agree with the determiner; without an active gender feature in the SS, there is no way to account for double-definiteness marking in Norwegian. As such, it is not possible for a donor language to supply the syntax for an intra-word codeswitch.

If, over the course of a sentential derivation, the dominant contexts switch (not in rapid oscillation but a single alternation), then it is possible to construct a syntactic frame with elements from two languages—however, as contexts are not oscillating host and

donor languages are not in competition. This will result in intra-sentential codeswitching as only one language is conceptually triggered at a time.

Ultimately, the MOGUL framework places no overt restrictions on codeswitching based on the cognitive architecture or operations. Even Poplack's illicit examples of codeswitching (e.g. *eat-iendo*) are theoretically possible. From a performance perspective, the term *eat-iendo* will be interpretable by a Spanish-English bilingual as the concept *eat* + duration/continuation even if they deem it to be ungrammatical. Assuming MacSwan's assertion (2005)—that codeswitching is the product of a union between two grammars—the ungrammatical interpretation of *eat-iendo* is likely the result of a specific set of language mixing rules resulting from the union of the two grammars involved (i.e. American English & Puerto Rican Spanish). While I can only speculate as to the specific reason why this Puerto Rican community of language mixers deem *eat-iendo* to be ungrammatical, there are a number possible reasons. For example, tacit knowledge that the word 'eat' is an irregular verb in English may create barrier for co-indexation with regular Spanish suffixes. Or perhaps, as Spanish is a Romance language, this community of language mixers may have special rules dealing with English roots depending on whether they are Germanic or Latinate in origin. Regardless of why the term is deemed ungrammatical or how strange this term may sound to a Spanish/English bilingual, the MOGUL architecture ascribes meaning to all input/intake; given the decompositional nature of *eat-iendo*, the mind/brain will correctly interpret this term even if the speaker/listener thinks it wrong and would never produce the term themselves²⁸. This notion, that codeswitching is the blending of language varieties, is evident when examining intra-word codeswitching in various populations (e.g. Spanish-German intra-word codeswitching in Gonzalez-Vilbazo & Lopez, 2011).

However, this does not mean grammaticality judgments involving codeswitching are *laissez faire*. Accepting MacSwan's assertion (i.e. codeswitching = $G_x \cup G_y$; see chapter 3 for details), any rules or restrictions that govern codeswitching are the unique product of the languages which are being mixed. So, the formation of the codeswitch *den*

²⁸ If a bilingual believes a term to be ungrammatical it will be co-indexed to a negative value representation which will impede the co-indexed chain's ability to compete from lexical selection against conceptually equivalent terms with higher value representations.

field-a is governed by the union of an English grammar and a Norwegian grammar. While there is no definitive way to predict what the outcome of a union between two grammars will look like, it is possible to formulate a specific set of rules/restrictions for the mixing of two language-varieties based on observations from empirical evidence. For example, in the English/Norwegian codeswitching data presented in this thesis, all instances of intra-word codeswitching adhere to Norwegian double-definite marking (i.e. where Def, Gen & Num features on the determiner Agree with the affix). As such, a researcher investigating codeswitching in the CANS corpus (the Corpus of American Norwegian Speech) may predict that a codeswitch like *den field-er* will be considered ungrammatical because the determiner *den* does not agree with Norwegian suffix *-er* in number, definiteness or gender. This means that while predictions can be made regarding the grammaticality of codeswitches, these predictions are specific to a set of language-varieties and are not universally applicable.

4.3.1 Discussion: From Codeswitching to Loanwords

An unexpected consequence of this MOGUL account of intra-word codeswitching is a natural account of loanword development. The story of intra-word codeswitching, as discussed above, accounts for how representations associated to different language contexts can appear in the same derivation. Once a representational chain with mixed languages elements is constructed, APT (Acquisition by Processing Theory) provides a pathway for loanword integration.

When a novel representational chain is constructed, module-specific representations are co-indexed together; in a mixed language circumstance, where there is an oscillation between two dominant languages, the newly formed representational chain will be co-indexed to both an L_x and an L_y general language representation in the CS—this allows for L_x features to be associated with an L_y context and visa versa. Each time the new representational chain becomes active, the associations between chained representations are strengthened and the chain's resting level of activation increases. Over time, as the L_x PS representation is repeatedly co-indexed to an L_y general language representation (i.e. GLR), the mixed language element becomes permanently associated to the L_y context. When this occurs, the mixed language elements are said to

be integrated into the host language; what was originally an Lx PS representation is now an L(x)y representation which can now be triggered via an Ly cognitive context.

Phonological integration is a key requirement of loanword formation. Once the L(x)y PS feature bundle has been co-indexed to a Ly GLR they will compete with Ly PS features for selection. However, the new L(x)y PS representation is no longer receiving an activation boost from the Lx GLR and will have a relatively low resting level of activation, as a result, over time, Ly PS feature bundles (i.e. phonemes) will begin to replace L(x)y feature bundles (i.e. phonemes) during regular competition. This occurs because the new L(x)y feature bundles representing ‘foreign’ phonemes from Lx may contain ‘alien’ features which have no/few co-indexed representations when compared to Ly features—ultimately, common Ly features will be more active than new L(x)y features in an Ly context.

Consider an example where a French speaker creates an English/French intra-word codeswitch which contains the phoneme [θ] with English as the donor language. The intra-word codeswitch is co-indexed to a French GLR; upon repeated usage of the L(x)y (x-English; y-French) representation the strength of the L(x)y - GLR co-indexation will increase which allows the new L(x)y representation to be triggered by a French context. However, when a French context triggers a L(x)y representation, the L(x)y features strongly associated with English must compete without the benefit of a GLR activation boost—this means when the English phoneme [θ] competes with French [t] for selection, the French [t] will win. Upon repeated activations, the strength of the co-indexation between the French [t] and the L(x)y representation will increase, while, conversely, the infrequent activation of English [θ] will cause its co-indexation with the French GLR to slowly attrite. The result is that, given time and repeated activations, an English codeswitch will be fully phonologically integrated into the French language and the codeswitch will become a loanword.

While the details of this analysis need to be fleshed out, examples of this type of loanword integration are found in LaCharité & Paradis (2005) who investigate the phonological development of English loanwords in Mexican Spanish; specifically, they are interested in vowel change (i.e. $i \rightarrow i$; $u \rightarrow u$).

Table 4: English Loanwords in Mexican Spanish

	English	IPA	Mexican Spanish
[ɪ]	<i>Building</i>	[bɪldɪŋ]	[bɪldiŋ]
	<i>Busy</i>	[bɪzi]	[bisi]
[ʊ]	<i>Cook</i>	[kʊk]	[kuk]
	<i>Plywood</i>	[plajwʊd]	[plajwud]

It is worth noting that a MOGUL account of loanword integration appears to support LaCharité & Paradis' analysis (2005) that loanwords are phonologically integrated into their host languages. However, the exact details of how loanwords are phonologically integrated in MOGUL are beyond the scope of this thesis and are left to future research.

Section 4.4.0 Chapter Summary:

This chapter presented an account of intra-word codeswitching, arguing that codeswitching is a natural consequence of the MOGUL framework. This account of codeswitching relies on key elements from DM, notably the decompositional nature of lexical items and the late insertion of Vocabulary items into a fully constructed syntactic frame. While this account of intra-word codeswitching adopts MacSwan's assertion on codeswitching (i.e. nothing constrains codeswitching apart from the grammars involved), this thesis rejects his lexicalist assumptions (i.e. the lexical integrity principle) and their implications (i.e. the PS Disjunction Theorem).

The cognitive processes which allow for language mixing to occur is exemplified via a detailed account of the English-Norwegian codeswitch *den field-a*. This example demonstrates how oscillating dominant contexts and goal representations influence activation levels of linguistic representations in the language core. Notably, while oscillating contexts allowed for representations associated with different languages to compete for selection, it is co-indexation to the goal representation which causes the PS representation from the donor language to become most active in the PS. The result is the formation of a $[PS_{Lx} + SS_{Ly} + CS_{Lx}/CS_{Ly}]$ representational chain which manifests as

an intra-word codeswitch. Notably, in this account of intra-word codeswitching, language mixing is a natural product of standard MOGUL operations and processes.

Finally, this account of intra-word codeswitching makes empirical predictions. Most notably, this thesis predicts that in order for intra-word codeswitching to occur the host language must supply the syntactic frame while donor language supplies the PS representation for the switched element. It is not possible to produce an intra-word codeswitch where the donor language supplies the syntactic frame and the host language supplies PS representations—intra-word codeswitching of this type will be blocked either by insufficient activation levels (e.g. *the field* – with Norwegian gender feature) or by the subset principle (e.g. *den felta* – with English SS features/no gender feature).

Chapter 5: Conclusion

5.0.0. The Future of MOGUL

MOGUL is a new and ambitious framework that is still in its formative years of development. Since its introduction in 2004, the framework has expanded to include numerous theories and models which help explain various linguistic phenomena while developing a larger picture of language and cognition. For example, the notion of conceptual triggering was introduced by SS&T in 2014 to provide a MOGUL alternative to language tags (SS&T 2004), while value representations were first discussed in SS&T (2016) and functioned to provide motivation for language selection. While these concepts were originally introduced help account for specific linguistic processes, they have proven useful when investigating a wide array of linguistic phenomenon—notably, language mixing.

As MOGUL continues to grow, researchers are testing the efficacy of various linguistic and cognitive models within the framework. Some attempts have been more successful than others—during a brief discussion on language mixing, SS&T (2014) note that MacSwan’s lexicalist approach to codeswitching shows promise but leave the details for future investigation. However, as mentioned in chapter 3, the PS Disjunction Theorem (see section 3.2.3. for details) makes MacSwan’s approach to language mixing incompatible with intra-word codeswitching. Additionally, the molecular view of a word in MOGUL conflicts with a lexicalist approach to the storage of lexical items (i.e. the homogenous storage of syntactic, phonological and semantic information). This means that MacSwan’s lexicalist model is not only unsatisfactory to account for intra-word language mixing but is incompatible with MOGUL.

Enter Distributed Morphology. DM, just like MOGUL, presents a molecular view of a word where syntactic, phonological and semantic information are distributed across multiple lists (i.e. modules). This division of syntactic, phonological and semantic information allows for elements associated with different languages to manifest in a single word. While this allows for intra-word codeswitching, this analysis is also consistent with the greater MOGUL framework and can be exploited to account for numerous linguistic phenomenon unrelated to language mixing. Notably, by situating a DM account of I-language within the MOGUL, the framework may be able to adopt, at

least as a starting point, a number of DM analyses for various other linguistic phenomenon (e.g. the omission of L2 inflectional morphemes). While this thesis argues for the adoption of DM into the MOGUL framework, further investigation is required to test the efficacy of DM to account for additional linguistic phenomenon within the MOGUL framework.

Indeed, as pointed out by SS&T, there is still much work that needs to be done in the MOGUL framework. While some theories, like DM, require additional support and testing, other aspects of MOGUL remain under construction. For example, as it stands now, the CS (Conceptual Structure) in MOGUL is the source of some confusion—this module is said to be responsible for representing theta roles, GLRs (General Language Representations) and goal representations. How exactly this module represents these disparate features using a single set of primitives is not completely understood. As research continues in MOGUL, puzzling elements of framework, like the CS, will become more clearly defined. The exact shape and nature of CS primitives is an open question for MOGUL researchers.

However, this muddy picture of the CS may be seen as stepping-stone of sorts; as researchers continue to build the MOGUL framework, previous assumptions may need to be revisited and revised in light of new evidence. For example, as noted in chapter 2.5.1, SS&T presented two possibilities for representing goals in MOGUL: goals can either be part of a complex CS representation, or they may belong to their own ‘Goal Module’; as there is currently no evidence to support one option over the other, SS&T opted for the more parsimonious assumption and placed goal representations inside the CS. Nonetheless, future research may, hypothetically, provide counter-evidence and suggest that goal representations belong to their own module which will require a re-analysis of goals in MOGUL. Regardless of what insights the future holds, the treatment of goal representations as complex CS representations in this thesis is consistent with the greater MOGUL architecture and has allowed this thesis to build arguments regarding the interactions between goal and value representations. As research into the MOGUL framework progresses and empirical evidence which supports a specific theory over an alternative is discovered, a clearer, more cogent model of the CS and goal representations may be developed. Given the scope and ambition of the MOGUL framework further

research is needed to support and corroborate the theories adopted by MOGUL. MOGUL is, very much, a work in progress.

5.0.1. Future Directions: The Quantum Metaphor

The metaphors that are used to discuss language and the brain have a critical role in shaping how we actually think about language and the brain (Lakoff, 1980). For example, the term ‘module’ was introduced in chapter 2 to describe a specific set of processes which process a specific type of information. However, the term ‘module’ carries with it certain metaphorical baggage—notably, the notion that a ‘module’ is a bounded entity (i.e. an object with shape and edges) which resides in a physical location as opposed to the decentralized series of processes described in MOGUL; these misleading connotations are an unavoidable consequence of imprecise or imperfect language used to describe abstract phenomena.

Researchers investigating language and the mind/brain have been searching for new metaphors to describe the complex and abstract operations involved in language processing. Of particular interest to the MOGUL framework is the ‘Quantum metaphor’ introduced by Libben (2017). Libben claims that the adoption of certain metaphors from quantum physics—notably wave-particle duality (i.e. light is both a wave & a particle when observed) and superposition—can help linguists understand the complexities of word formation in the mind/brain.

In quantum physics the concept of superposition is often illustrated with a famous example known as Schrodinger’s Cat: in a rather cruel thought experiment, a cat is placed inside a sound proof box with a portion of poison food. When the box is closed, the cat is said to exist in a superposition where it may be considered both alive and dead until the box is opened and the cat is observed, in which case the superposition collapses and the cat is in one of two possible states—either alive or dead.

To briefly explore this metaphor in MOGUL, consider the notion of oscillating contexts and dominant languages. As described in chapter 4, the dominant cognitive context conceptually triggers a specific set of linguistic representations. In the case of codeswitching, two contexts are more or less in balance and there is a rapid oscillation between these contexts in real time over the course of an utterance, which results in representations from two languages competing for insertion into the derivation.

Alternatively, this process can also be understood in terms of superpositions. In a grossly simplified account, dominant contexts (i.e. oscillating contexts) can be said to co-exist in a superposition where neither is realized until it is observed, in which case the superposition collapses into a single dominant context. Crucially, the adoption of a quantum metaphor to describe dominant contexts has consequences for how we think about the subsequent processes in language selection. Notably, the metaphor that dominant contexts exist in a superposition excludes the possibility of competition between dominant contexts, since states co-existing in superposition do not compete. Abandoning the ‘competition’ metaphor may fundamentally change the account of intra-word codeswitching presented in this thesis. Needless to say, the fallout from adopting a quantum account of mental processing has serious repercussions for how researchers think about operations in the mind/brain. Nonetheless, as pointed out by Libben, the quantum metaphor may also shed new light on previously unseen or misunderstood linguistic phenomenon. The consistent application of the quantum metaphor to the MOGUL framework may reveal promising new insights into language processing in the mind/brain.

5.1.0 Final Summary & Conclusion

The MOGUL framework was designed by SS&T to act as a platform to bring together language-orientated researchers and academics from a wide variety disciplines; allowing them to integrate their work into a unified model of the mind/brain—SS&T describe MOGUL as a ‘super-convergence zone’ (2014, p. 360). Following this call-to-arms, this thesis attempted to use the MOGUL platform to bridge the gap between language competence and production in regards to intra-word codeswitching. To this end, a DM, competence-based model of language mixing, was situated inside of the performance-based MOGUL framework—DM provided an account of syntactic relations while MOGUL provided an account of context and goal representations exploited during the production of codeswitches. By doing so, the vision of language presented in this thesis was able to use a model of linguistic competence to inform a model linguistic performance. True to the interdisciplinary nature of MOGUL, the phenomenon of language mixing was used to form a bridge between generative grammar and cognitive science.

One aim of this thesis was to demonstrate that language mixing is a natural consequence of the MOGUL framework. In the MOGUL framework, representational chains are constructed from the co-indexation of module-specific feature bundles (e.g. PS + SS + CS). These chains constitute a molecular view of word (see section 2.5.1), which, ultimately, allow for modular specific features bundles associated with different languages to come together to form intra-word codeswitches. These intra-word codeswitches are the natural consequence of an oscillating cognitive context interacting with goal representations; the resulting configuration of cognitive apparatuses constitutes a ‘bilingual communicative mode’. When this communicative mode is engaged, two languages are conceptually triggered and all representations co-indexed to either GLR receive an activation boost; as such, module-specific representations associated different languages are in competition for selection. Crucially, goal representations are co-indexed to any representation which helps satisfy the goal which, in turn, increases the activation level of said representation; when the goal can be satisfied by the production of an intra-word codeswitch, the activation level of PS representations from the donor language receive an additional activation boost and becomes the most active PS representation. A representational chain is then formed from the most active content in each module; the resulting chain will contain a SS representation from L_x, a PS representation from L_y and CS representation from both L_x and L_y—this is an intra-word codeswitch. Notably, there are no special processes or mechanisms evoked to account for codeswitching; language mixing is the natural product of standard MOGUL operations and processes. The account of intra-word codeswitching described here follows the claim by Sharwood-Smith & Truscott who state that “a theory of codeswitching should, ideally, not be a theory of codeswitching” (2016 p.1). Indeed, the picture of the multilingual mind presented in the MOGUL framework has been central to the claim that codeswitching is the natural consequence of bilingualism (or multilingualism).

The chief achievement of this thesis was to propose a way to encode a DM competence theory within a MOGUL performance theory, thus allowing MOGUL to account for intra-word codeswitching. Central to this account of intra-word codeswitching is the insertion of DM Vocabulary Items into a fully constructed syntactic frame, where the syntactic features are supplied from the host language and the donor

language provides the switched Vocabulary item. Notably, this approach to intra-word codeswitching offers empirical predictions. It was argued, in section 4.3.0, that in order for intra-word codeswitching to occur the host language must provide the syntactic frame while the switched Vocabulary item must come from the donor language.

Hypothetically, an attempt to construct an intra-word codeswitch where the donor language provides the syntax and the Vocabulary items come from the host language will fail because: a) PS representations (e.g. Vocabulary Items) from the host language which are not co-index to active SS features would not achieve sufficient levels of activation, or b) Vocabulary Items that are co-indexed to SS features which are not active would violate the Subset Principle. As such, the account of intra-word codeswitching presented here, predicts that intra-word codeswitching may not occur if the syntactic frame is from the donor language and the Vocabulary items are from the host language.

As a final note, the story of language mixing in the MOGUL framework is far from complete. As MOGUL continues to develop its account of representation and processing in the multilingual mind, the phenomenon of language mixing will play a central role in accounting for how a bilingual organizes two distinct language systems. As interdisciplinary investigations into language mixing continue, intra-word codeswitching will continue to offer a fascinating window into language representation and processing in the multilingual mind.

Bibliography

- Abutalebi, J., Annoni, J.-M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., ... Khateb, A. (2008). Language control and lexical competition in bilinguals: an event-related fMRI study. *Cerebral Cortex*, *18*(7), 1496–505. doi:10.1093/cercor/bhm182
- Abutalebi, J., Della Rosa, P. a, Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(3), 905–11. doi:10.1016/j.cortex.2012.08.018
- Amaral, L., & Roeper, T. (2014). Multiple Grammars and second language representation. *Second Language Research*, *30*(1), 3-36.
- Arad, M. (2003). Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory*, *21*(4), 737-778.
- Alexiadou, A., Lohndal, T., Åfarli, T. A., & Grimstad, M. B. (2015). Language mixing: a distributed morphology approach. In T. Bui, & D. Özyildiz (Eds.), *Proceedings of NELS* (Vol. 45, pp. 25-38).
- Belazi, H., Rubin, E., & Toribio, A. (1994). Code Switching and X-Bar Theory: The Functional Head Constraint, *Linguistic Inquiry*, *25* (2).
- Blanco-Elorrieta, E., & Pylkkanen, L. (2015). Brain bases of language selection: MEG evidence from Arabic-English bilingual language production. *Frontiers in Human Neuroscience*, *9*(27), 1–19. <http://doi.org/10.3389/fnhum.2015.00027>
- Blanco-Elorrieta, E., & Pylkkanen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *Journal of Neuroscience*, *36*(2), 290–301. <http://doi.org/10.1523/JNEUROSCI.2597-15.2016>
- Bliss, H. (2005). Constructing dual number in Hopi. In Proceedings of the 2004 Canadian linguistics association annual conference (pp. 1-12).
- Borer, H., & Wexler, K. (1987). The maturation of syntax. In *Parameter setting*. Springer Netherlands, 123-172
- Brensan, J. & Mchombo, S. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, *13*(2), 181-254

- Bromberger, S., & Halle, M. (1989). Why phonology is different. *Linguistic inquiry*, 20(1), 51-70.
- Burkholder, M., Mathieu, E., & Sabourin, L. (2017) The role of gender in mixed language nominal phrases: Insights from Distributed Morphology. GALSA 14, Southampton, UK.
- Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition* (Vol. 25). John Benjamins Publishing.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- Chomsky, N., & Lasnik, H. (1993). Principles and parameters theory. *Syntax: An international handbook of contemporary research*, 1, 505-569.
- Chomsky, N. (1995). *The minimalist program* (Vol. 28). Cambridge, MA: MIT press.
- Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain?. *Nature Reviews Neuroscience*, 15(5), 336-345.
- Dijkstra, T., Haverkort, M. (2004) APT: Costs and benefits of a hybrid model. *Bilingualism: Language and Cognition*, 7(01), 26-28.
- Embick, D., & Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. *The Oxford handbook of linguistic interfaces*, 289-324.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- Foster, P. S., Hubbard, T., Campbell, R. W., Poole, J., Pridmore, M., Bell, C., & Harrison, D. W. (2017). Spreading activation in emotional memory networks and the cumulative effects of somatic markers. *Brain informatics*, 4(2), 85-93.
- Garbin, G., Costa, a, Sanjuan, a, Forn, C., Rodriguez-Pujadas, a, Ventura, N., ... Avila, C. (2011). Neural bases of language switching in high and early proficient bilinguals. *Brain and Language*, 119(3), 129–35. doi:10.1016/j.bandl.2011.03.011
- Gonzalez-Vilbazo, K., & Lopez, L. (2011). Some properties of light verbs in code-switching. *Lingua*, 121, 832–850. <http://doi.org/doi:10.1016/j.lingua.2010.11.011>
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1(2), 67-81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(July 2015), 1–16. <http://doi.org/10.1080/20445911.2013.796377>

- Grimstad, M. B., Lohndal, T., & Åfarli, T. A. (2015). Language mixing and exoskeletal theory: A case study of word-internal mixing in American Norwegian.
- Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing*, 122.
- Halle, M. (2000) "Distributed morphology: Impoverishment and fission." Amsterdam Studies in the Theory and History of Linguistic Science Series 4: 125-150.
- Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. *MIT working papers in linguistics*, 21(275), 88.
- Harley, H. (2014). On the identity of roots. *Theoretical linguistics*, 40(3/4), 225-276.
- Harley, H., & Noyer, R. (2000). Formal versus encyclopedic properties of vocabulary: Evidence from nominalizations. *The lexicon-encyclopedia interface*, 349374.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26(2), 210-231.
- Haugen, J. D., & Siddiqi, D. (2013). Roots and the derivation. *Linguistic Inquiry*, 44(3), 493-517.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative?. *Linguistics*, 39(6; ISSU 376), 1041-1070.
- Jakobson, R., Fant, G., & Halle, M. (1951). Preliminaries to speech analysis. The distinctive features and their correlates.
- Jackendoff, R. (1997). *The architecture of the language faculty* (No. 28). MIT Press.
- Jackendoff, R. (2003). Précis of Foundations of Language : Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*, 26(6), 651–707.
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science*, 4(1), 71-115.
- Kawase, S., Hannah, B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *The Journal of the Acoustical Society of America*, 136(3), 1352-1362.
- Krämer, M. (2012). *Underlying representations*. Cambridge University Press.
- Libben, G. (2017). The quantum metaphor and the organization of words in the mind. *Journal of Cultural Cognitive Science*, 1, 49-55.

- Lohndal, T. (2013). Generative grammar and language mixing. *Theoretical Linguistics*, 39(3-4), 215-224.
- Paradis, C., & Lacharité, D. (1997). Preservation and minimality in loanword. *Journal of Linguistics*, 33(2), 379-430.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: language and cognition*, 3(1), 37-54.
- MacSwan, J. (2005). Précis of a minimalist approach to intrasentential code switching. *Italian Journal of Linguistics*, 17(1), 55.
- MacSwan, J. (2014). *A minimalist approach to intrasentential code switching*. Routledge.
- Marantz, A. (1995). Cat as a phrasal idiom: Consequences of late insertion in Distributed Morphology. ms., MIT.
- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics*, 4(2), 14.
- Myers-Scotton, C. (2000). Explaining the role of norms and rationality in codeswitching. *Journal of Pragmatics*, 32(9), 1259-1271. [http://doi.org/10.1016/S0378-2166\(99\)00099-5](http://doi.org/10.1016/S0378-2166(99)00099-5)
- Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development*, 13(1-2), 19-39. <http://doi.org/10.1080/01434632.1992.9994481>
- Peperkamp, S. (2003). Phonological acquisition: Recent attainments and new challenges. *Language and Speech*, 46(2-3), 87-113.
- Peperkamp, S. (2004). A psycholinguistic theory of loanword adaptations. In *Annual Meeting of the Berkeley Linguistics Society* 30(1), 341-352).
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en espa??ol: Toward a typology of codeswitching. *Linguistics*, 18(7-8), 581-618. <http://doi.org/10.1515/ling.1980.18.7-8.581>
- Poplack, S. (1988). Contrasting patterns of code-switching in two communities. *Codeswitching: Anthropological and sociolinguistic perspectives*, 215, 44.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26(1), 47-104. <http://doi.org/10.1515/ling.1988.26.1.47>

- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1), 73-193.
- Roeper, T. (1999). Universal bilingualism. *Bilingualism: language and cognition*, 2(3), 169-186.
- Roeper, T. (2004) Can the APT be a theory of Multiple Grammars and a minimal theory of parsing? *Bilingualism: Language and Cognition*, 7(01), 40-42.
- Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1987). *Parallel distributed processing* (Vol. 1, p. 184). Cambridge, MA, USA:: MIT press.
- Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language variation and change*, 2(1), 71-101.
- Smith, M. S., & Truscott, J. (2014). *The multilingual mind: A modular processing perspective*. Cambridge University Press.
- Truscott, J. (2004). How APT is your theory: Present status and future prospects. *Bilingualism: Language and Cognition*, 7(01), 43-47.
- Truscott, J. (2014). Consciousness and second language learning. *Multilingual Matters*.
- Truscott, J., & Sharwood-Smith, M. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, 7(1), 1-20.
- Truscott, J., & Sharwood-Smith, M. (2016). Representation, Processing and Code-switching. *Bilingualism: Language and Cognition*, 1-14.
- Van Dijk, T. A. (1997). Cognitive context models and discourse. *Language structure, discourse and the access to consciousness*, 189-226.