Explorations in Second Language Acquisition and Processing

Explorations in Second Language Acquisition and Processing

Edited by

Roumyana Slabakova, James Corbet, Laura Dominguez, Amber Dudley and Amy Wallington

Cambridge Scholars Publishing



Explorations in Second Language Acquisition and Processing

Edited by Roumyana Slabakova, James Corbet, Laura Dominguez, Amber Dudley and Amy Wallington

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data A catalogue record for this book is available from the British Library

Copyright © 2019 by Roumyana Slabakova, James Corbet, Laura Dominguez, Amber Dudley, Amy Wallington and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-2776-X ISBN (13): 978-1-5275-2776-8

TABLE OF CONTENTS

Introduction
Part I: Plenary Talk
Chapter One
Part II: The acquisition of subject and object pronouns
Chapter Two
Chapter Three
Chapter Four
Chapter Five
Chapter Six

Part III: The Syntax-Discourse and the Syntax-Semantics Interfaces

The Role of Input in the Acquisition of English Generics by L1 Najdi Arabic Speakers May Abumelha

Part IV: Acquisition of Syntax

Chapter Nine	154
Investigating Co-activation of L1 Syntax during Processing of <i>Wh</i> -questions: Eyetracking Evidence from L1 German–L2 English	
Tom Rankin, Theres Grüter and Holger Hopp	
Chapter Ten	171
The L2 Acquisition of English VP-ellipsis and Adverb Placement Kholoud Al-Thubaiti	
Chapter Eleven	185
Anaphor Reconstruction in L1 and L2 Japanese Relative Clauses	
Yunchuan Chen and Shin Fukuda	
Chapter Twelve	197
Grammar Competition in Second Language Acquisition:	
The Role of V2 in wh + Particle Constructions	
Leah S. Bauke	
Chapter Thirteen	208
Competing Factors in the Acceptability of Italian Embedded Wh-questic	ons
Stefano Quaglia, Tanja Kupisch and Anika Lloyd-Smith	

vi

Explorations in Second Language Acquisition and Processing	vii

Part V: Acquisition of Functional Morphology

Chapter Fourteen	222
The Role of Gender in Mixed-language Nominal Phrases:	
Insights from Distributed Morphology	
Michele Burkholder, Éric Mathieu and Laura Sabourin	
Chapter Fifteen	238
Nonword Pluralization: Evidence for an Advantage of Bilingualism	
in Albanian–Italian and Romanian–Italian Bilingual Children	
Chiara Melloni, Maria Vender and Denis Delfitto	

INTRODUCTION

This volume springs from the 14th edition of the Generative Approaches to Second Language Acquisition conference. GASLA 14 was held at the University of Southampton from 7 to 9 April 2017. It was the first European edition of this conference, which had been held in North America before this.

The oral presentations and the posters presented at the conference were selected from the one hundred submitted abstracts, which went through a rigorous and anonymous review process. After the conference, presenters were invited to submit their work to this proceedings volume, which you are now holding in your hands. We are grateful to all authors for sharing their work.

Several themes and directions of research were visible at the conference and we have organized the different parts of the volume around them. The plenary by **Michael Sharwood Smith** develops further his ground-breaking theoretical model on bilingual representation and processing, providing the cornerstone of the volume. The chapter is particularly valuable to generative linguists for the clarity of the example of gender representation and processing, which runs through all the main sections.

The acquisition of subject and object pronouns was another topic of enhanced interest at the conference. The two chapters on L2 Portuguese by **Lobo, Madeira and Silva** and by **Fiéis and Madeira** look at native language transfer effects in choosing antecedents for null and overt pronouns, and at the interpretation of pronouns and clitics by Chinese learners. The next two chapters by **Kraš and Milicevic Petrović** and by **Milicevic Petrović, Kraš and Lisica** take up anaphora resolution in the competence of simultaneous Croatian–Italian bilinguals and translators. The bilingual participants were comparable to monolinguals in their choices, suggesting that difficulties are encountered only when languages mismatch. Neither the experienced nor the trainee translators demonstrated L1 attrition for the property. The final chapter in this part is by **Kubota**, who discusses referential interpretation and how that interplays with cognitive control in Japanese–English bilingual children.

The next two chapters also engage with interpretation, but in constructions which can be fully appreciated only at the interfaces between

interface. The chapter by **Teixeira** takes up locative inversion while the chapter by **Abumelha** considers generic constructions in L2 English. Teixeira asks if the syntax–discourse interface is a locus of permanent optionality and answers in the affirmative, adding some modulating factors. Abumelha uses linguistics-informed methods to teach generics under explicit and implicit instruction conditions.

Acquisition of syntax is the focus of the next part of the volume. Rankin, Grüter and Hopp investigate co-activation of the native syntax during processing of *wh*-questions in L1 German–L2 English interlanguage, using the evetracking method. Al-Thubaiti discusses verb phrase ellipsis and adverb placement in the grammar of Arabic-native learners of English, both arguably dependent on an uninterpretable feature. L1 and L2 relative clauses and the process of reconstruction are the topic of the chapter by Chen and Fukuda. The authors argue that, for Chinesenative learners of Japanese, this is a poverty of the stimulus property and show that it is indeed acquired successfully. Bauke is interested in the grammar competition between V2 and non-V2 options in L2 English, using the wh + particle construction to tackle this matter. The chapter by Quaglia, Kupisch and Lloyd-Smith investigates embedded whquestions, a construction that displays a high degree of variation in monolingual Italian. They look at various factors that can explain crosslinguistic influence in heritage and monolingual speakers of Italian.

The final part of the volume contains two chapters on functional morphology. Based on the Distributed Morphology model, **Burkholder**, **Mathieu and Sabourin** provide a theoretical proposal and experimental evidence for the role of gender in mixed-language nominal phrases. **Vender**, **Delfitto**, **Mantione and Melloni** are interested in whether Albanian–Italian and Romanian–Italian bilingual children show any bilingual advantage when asked to inflect real words and nonwords for the plural. They show that the bilingual participants indeed demonstrate this advantage.

The breadth and variety of the topics in this volume's research is a testament to the vitality and rigor of generative second language acquisition. At the same time, the chapters make it clear that grammatical representation, processing and context have to be engaged together, in order to elucidate the process in all its complexity and richness.

The Editors

PART I:

PLENARY TALK

CHAPTER ONE

REPRESENTING REPRESENTATION: THREE PERSPECTIVES AND THE CASE OF GRAMMATICAL GENDER¹

MICHAEL SHARWOOD SMITH

1. Overview

In this chapter, I will present a processing-based working model of the mind, based on research findings across a range of disciplines within cognitive science. The inclusion of processing considerations should not obscure the fact that representational and processing explanations are integrated within this model, or more properly, within this theoretical framework. This makes it an extension of theoretical linguistic explanations for changes in the way a language is represented in the mind of an individual. It also runs counter to the current and, in present terms, entirely misguided tendency to see representations and processing routines as entirely separate phenomena. Where research deals with acquisition in real time, as is the case with developmental linguistics, only an integrated view makes sense. A representation existing in the mind of a specific individual engaged in language-related activity is a particular combination of structural and processing properties. These can change together over time and in different ways: you cannot consider one without considering the other.

The role of language is interpreted, in line with the generative enterprise, as being dependent on a uniquely human, biologically endowed linguistic ability. Language ability in its broadest sense depends on this core ability but is actually much more extensive, involving many parts of the mind that have other unrelated functions. Any unifying framework that encompasses all these aspects will need to incorporate much more than an

¹ Grateful acknowledgments are due to the editors, my anonymous reviewers and a productive exchange of views with John Truscott.

abstract account of linguistic structure divorced from time and space considerations. The manner in which its theoretical insights are formulated out for internal theoretical purposes will not be a reliable and complete guide when working out the nature of those mechanisms responsible for online processing, storage and development. The underlying aim is, accordingly, to integrate theoretical linguistic accounts with current explanations of how the mind processes and stores mental representations of any kind. This has also to be done in a way that is in tune with and can supplement work in current neuroscience.

A platform like this has arguably not been available to researchers thus far and, not surprisingly, researchers have become so accustomed to doing without one that they seem to have lost any sense of really needing one in the first place. I would argue strongly that such a conceptual platform, providing as it does a much clearer working model of the mind as a whole than one based on vague assumptions, should not be regarded as a luxury extra or perhaps just something for the future but rather as a dire necessity. It does require the abandonment of more locally-based frameworks for the guidance of research. In fact, it depends on them. At the same time, separate hypotheses and theories developed and tested using terminology and techniques that facilitate empirical work only within one individual research area do little to promote a combined view of what they all mean for our understanding of the mind. The tradition of studying separately linguistic representations, on the one hand, and on-line language processing, on the other, provides a prime example of academic apartheid that is not in the long term helpful for those interested in accounting for development which requires both perspectives to be combined.

After very briefly discussing the basic features of this framework, I will go on to show, in terms of the framework, how language cognition fits with cognition in general. This will include accounting for how two or more languages can be accommodated within the same mind and providing precise definitions for some crucial concepts that are often avoided or mentioned in imprecise terms. As suggested by the chapter subtitle, the life cycle of a representation, the chapter will conclude with a brief implementation. This will look at gender. Implementations of the framework can of course differ: alternative explanations can be proposed so this will just be an example to illustrate how the framework can be exploited. Grammatical gender will in fact form a thread running through the whole discussion.

Chapter One

2. Nesting Frameworks

Every researcher in language acquisition must of necessity work with some mental model of how the mind is organised. The model may be largely implicit although there will be explicit, more elaborated parts, reflecting just those areas where the researcher in question works and therefore is most expert. An obvious example in the context of this volume is what generative linguistic theory provides, including empirical studies in particular aspects of grammar and in particular languages. In addition, there may be other, somewhat less elaborated areas which have proved to be of obvious and immediate relevance. either in providing methodological tools or theoretical insights or both. What might some of those aspects of mental organisation be that one might expect any researchers to be knowledgeable about, that is, apart from the abstract linguistic properties of a given language system? Clearly, they would have to be issues concerning linguistic development in real time. Here is a small rough-and-ready checklist containing just a handful of the fundamental questions that would require coherent and detailed answers:

- 1. How do you personally imagine the way the mind instantiates and manipulates representations in real time?
- 2. What is working memory and how does it work?
- 3. "Interface" should be a familiar commonly used concept from the generative linguistic literature: how exactly do you think an interface works in real time to shed light on experimental results in studies of
 - a. on-line processing?
 - b. development over time?

My guess is that few researchers would like to be put on the spot and asked to produce detailed answers to these questions on the spur of the moment. The reason for that is twofold. First, given the current state of our knowledge in the relevant areas of cognitive sciences, there is very little consensus about any of these issues. Secondly, most researchers simply do not have the time available to acquaint themselves with the relevant research fields and review all the available options and will tend to plump for what seems to be the most accessible and long-established approach. One way out of this dilemma is to commit to interdisciplinary research projects, a norm in the hard sciences and increasingly popular now in cognitive science. Another way, especially compatible with the first, is the main focus of this chapter, namely to look for and use a wide-scope framework in which coherent, explicit commitments are made about just the kind of questions listed above. "Wide-scope" here means that in principle a given currently used framework in one or more subdomains can be nested within the wide-scope framework. One positive outcome of this nesting of frameworks, apart from its potential to expand and refine explanations of data, should be synergies that enable the refinement of the frameworks themselves. By the end of this chapter, there should be answers to each of the question listed above, each of which will be precise and coherent. Whether they are right or even useful is of course an empirical question.

There should be no controversy about the need to take a wider perspective on local theoretical issues. Indeed, it has increasingly been acknowledged in the research literature: wishes for this need to be addressed typically take one or other of two forms:

- 1. the whole-mind perspective
- 2. the mind/brain perspective

Expressing the whole-mind perspective, Thierry, a neurolinguist, writes as follows: "The time has come, perhaps, to go beyond merely acknowledging that language is a core manifestation of the workings of the human mind and that it relates interactively to all aspects of thinking" (Thierry 2016, p. 690). The second, mind/brain perspective is expressed by Kroll, a psycholinguist, who looks in the other direction: "Understanding how different aspects of language processing will engage cognitive and neural processes will be crucial" (Kroll 2015, p. 33). Although not a necessary implication, I choose to interpret these sentiments as an implicit plea for a facilitating framework of some kind. Such a framework should specify, with much greater precision than has been customary hitherto, the basic psychological processes and mechanisms involved. The proviso is only that it should take account of theoretical views and empirical findings in cognitive- and neuro-scientific research. Of necessity it will leave open much to be debated and refined.

3. The Interactive Modular Mind

3.1 The Framework

The wider-scope framework to be used in this discussion is the *Modular Cognition Framework* (MCF). A brief comment on this name is appropriate at this juncture. The framework has been known more widely

Chapter One

as the MOGUL (Modular Online Growth and Use of Language) framework. Accordingly, in all discussions relating to language cognition as in this chapter, it can still be used interchangeably with MCF: MOGUL happens to be the instantiation of the framework that is used specifically to explain language-related phenomena. In order to situate language within the mind as a whole so as to account for general issues such as perception, memory, cognitive control, attention and consciousness, it had been necessary to widen the scope to take that into account: the MCF name more appropriately reflects that mind-wide perspective, which should be equally applicable to research on areas of cognition other than language.

I will now outline briefly the main features of MOGUL (MCF). The mind, like the brain, has a modular architecture. In broad terms this is fairly uncontroversial, the devil being in the detail. In MCF, this means that the mind is composed of a network of interacting expert systems, each of which has an identical basic design. This basic design is of course neurally instantiated in many different ways but a mind-based account abstracts away from these. The modular system can be seen as a collaborative network, coping with a myriad of different tasks in parallel and with the modules connected with other modules by means of interfaces. Its neural instantiation will also be a network of interacting systems but will naturally look quite different. Unlike the interfaces posited purely within the context of mainstream generative linguistic literature, the framework versions are processors that operate in real time. They generally mirror the way interfaces are described in Jackendoff's architecture of the language faculty (see, for example, Jackendoff 1987, 2002). Moreover, these interfaces, as just mentioned, are not limited to just those that connect up the language-specific systems to adjoining systems outside but include all the other connections between modules as well

Another important point is that, within this collaborative network of modular systems, no system can be described as "domain-general." This term has been frequently used as a convenient way of saying "not in the language module" or "not governed by principles of UG" and so the implications of the existence of a domain general system have not been found relevant. In MCF, there are two potential candidates for domain-general status. The first obvious one is the *conceptual system*, which in the human mind forms a central hub for many of the mind's operations and, although its neural underpinning involves a number of different brain

systems, the most striking one is the human prefrontal cortex.² The second candidate is a temporary phenomenon related to the generation of conscious experience. In neural terms, it is what arises when intense synchronised online activity occurs in the various perceptual memory stores: these perceptual stores are coactivated with all the other stores that are relevant to the current context but it is indeed just these perceptual stores that create the experience we call "awareness" in both humans and other species. In a sufficiently cognitively advanced species, humans being the prime and perhaps only example, whatever conceptual content becomes available to awareness can immediately form the basis for conscious analysis, reflection and planning. Awareness of particular properties of language input, coupled with a preoccupation with aspects of grammatical gender, provide one relevant example here. Whether or not these higher levels of conscious cognition are active as well, the intense collaboration between individual memory stores, creates a combined effect called "global working memory" (GWM). This is the MCF version of a much-debated notion called the "global workspace" (Truscott 2017, cf. Baars 1988, 1997, 2002, 2007; Dehaene and Changeux 2011; Engle 2002). However, both these two candidates fail to fit the idea of a central processor where all mental activity is supervised. The conceptual system, even though it often plays the role of a central hub where connections between different stores intersect, still cannot be treated as a controlling mind-within-a-mind. In other words, it is no homunculus. Rather, it conforms like any other module to the same basic working principles in the way it works internally and the way it works externally via its interfaces with other modules. Global working memory, the other candidate, does in some measure have the appearance of possessing a supervisory, decision-making capability (which may be illusory, see Libet 1993). However, being exclusively about various degrees of awareness, it excludes by definition the vast majority of mental processes that work below the level of awareness and therefore beyond its direct control: it is an Oval Office with the computers down, the doors locked and very few people available to respond to orders. To sum up, whatever we may think, no one system controls the mind, which is not hierarchical but rather "heterarchical"³, without a permanent central executive in place.

² The ramifications of this are much more fully discussed in Truscott and Sharwood Smith (in prep.).

³ In other words, also in terms of the framework, all modules are equal: there is no master module (McCulloch 1945, Sharwood Smith & Truscott 2014, p. 21).

Chapter One

3.2 Modules: The Basic Design

I will now spell out, for those less familiar with MOGUL, the basic features of any module in the framework although some points will be made that have not been regularly emphasised in earlier publications. The reader is encouraged to keep in mind two things, firstly the fact that in most of the processing (psycho)linguistic literature there is frequent mention of given linguistics constructions being "harder to process" or "easier to process" on the basis of relative response times and/or measures of accuracy and secondly, the absence of any clear account of how the processing mechanisms that produce these particular responses actually work. This may seem an unjustified or exaggerated claim because frequent references are made to working memory and to particular accounts of what that is. Apart from the fact that working memory research is a dynamic field with much controversy and a variety of theoretical approaches on offer, the details of how processing works and the commitment to one or other of these approaches are not a regular feature of discussion sections of psycholinguistic studies (see Sharwood Smith 2017a and other contributions in the same special issue of Second Language Research).

A modular system contains a processor and a store. The processor is driven by principles unique to that module. For example, the syntactic module is constrained by syntactic principles, the nature of which is defined in various ways in syntactic theory. These principles have the effect of ensuring that any syntactic structure or "representation"⁴ known or created by an individual conforms to what can be thought of as a syntactic *code*.

A processor works in real time so it activates and assembles representations in response to current processing demands, which will be elaborated shortly. If we can say, using familiar terminology, that any syntactic representation must conform to the principles of UG, in this case as instantiated in the syntactic processor, then effectively we can *say that any module has its own UG*.⁵ The special contribution of syntactic theory is precisely to shed light on the properties of the syntactic processor and

⁴ From now on the terms "representation" and the philosophically non-committal "structure" will be used interchangeably i.e., as synonyms.

⁵ This characterisation of human mental modules in general stands in stark contrast to statements by those who would like to see the whole notion of UG; i.e., innate grammatical principles, however minimally defined, as a dead duck. In other words, it puts specifically <u>linguistic</u> properties (*N*, *V*, *Agr*, *sonorant*, *syllable* etc.) into a mind-wide context, as examples of what is actually *characteristic of cognition as a whole*.

the syntactic representations it handles. In fact, the notion of a modulespecific code which is part of our biological endowment can be applied across the board. In this way any *visual* representation must conform to visual UG, or in present terms, the *visual code*, and this will reflect innate, universal human principles governing human visual cognition. Note also that visual processing and syntactic processing in the brain, as indeed any other kind of neural processing, are distinguished by unique neural patterns as well. In other words, this modular idea can work in neural terms as well, although obviously in quite dissimilar ways as it involves particular brain locations, neural patterns and pathways.

Turning now to the stores which house the structures of a given module, these structures include *primitive elements*, structural properties or features from which more complex representations can be assembled and which reside in the system from the start. For example, the visual system will have its own primitives to help build complex representations of visual input over the life time, making human vision different from vision in any other species. Theoretical linguists may think of a store as a syntactic lexicon or "syntacticon" (Borer 1984; Emonds 1985, 2000). However, responding to input over the lifetime, these primitive structural elements may be combined in various lawful ways to build new complex (syntactic or other) structures in the appropriate (memory) store. The primitives are already there at birth, ready to be used when required.⁶ Hence a store will contain not only the primitives but all these more complex structures that have been created. A psycholinguist will see stores as *memories* and indeed this is what they are, one for each module. A modular approach to memory is very much in line with current thinking about memory in psychology as well as neuroscience (see for example D'Esposito and Postle 2015; Erikson et al. 2015; Oberauer et al. 2016).

3.3 Representations and Neurons

For those who are not classical connectionists, the basic idea of a representation is a familiar term and is defined in various ways according to a given researcher's theoretical stance. The above description of structures in a store as either basic simple structural items (primitives) or combinations of those items forming more complex items should not be objectionable. Whatever the preferred ways of describing representations in a given theory are, it should be easy for most people to reconceptualise

⁶ Alternatively, primitives may be thought of as belonging to the processor and only deployed in the store when needed.

Chapter One

a representation as a network, either as very local networks of features expressing some basic structural category or as a combination of such micro-networks into a larger, more complex representation. The designation of the smallest items as "primitives", that is, structural items that are provided as part of our biological inheritance, is more controversial but should certainly not be objectionable to those of generative linguistics persuasion. Furthermore, the description of structural items of greater or less complexity as being subject to processes such as storage and (co)activation will not give psycholinguists any cause for complaint, provided that they do not hold to the view that the concept of a representation is just a convenient post-hoc description for what is actually a set of interconnected nodes that have no symbolic function.

Interestingly, neuroscientists who may not regularly employ the term "representation" in describing neural phenomena may nevertheless still find the concept attractive, important and potentially very useful, Accordingly, some have felt the need to spell out what a representation might be in neural terms. Antonio Damasio, for example, uses the term "dispositional representation" which he defines as "a potential pattern of neuron activity in small ensembles of neurons" which "may be distributed over a number of different locations in the cortex" (Damasio 1994, pp. 102-105). Joaqin Fuster uses the term "cognit." the definition of which seems to fit best with the kind of representation, that involves a set of connections between representations ranging across different systems so not just representations within a single store (Fuster 2006, 2007, 2008). Both see representations as assemblies of interconnected neurons and Fuster makes clear that it is, for him at least, the neuronal patterns that count: the same cognit may involve different neurons on a subsequent occasion and still be the same cognit. He also assumes that some cognits are innate.

In sum, it seems appropriate to note that if neuroscientists feel the need to spell out what they think a representation might be for them, then researchers working in linguistics, in this case generative linguistics, should be equally interested in exploring and even developing the intellectual interface between theoretical psychological and linguistic constructs, on the one hand, and the theoretical concepts used in brain research, on the other. For this to work effectively, one needs a platform for doing this, which researchers working in these different areas can use; in other words, an overarching conceptual framework.

3.4 Module Cooperation: Interfaces and Schemas

In the mind, we can distinguish between the (sensory) perceptual systems, on the one hand, and the modules involved in deeper level processing, on the other. The perpetual systems responsible, respectively, for visual, auditory, gustatory, olfactory and somatosensory representations⁷ (seeing, hearing, taste. smell and body sense) together form the portal for the initial cognitive processing of environmental input (the ring of modules in Fig.1–1).



Figure 1-1. The perceptual portal featuring five stores [only stores and not their processors are displayed]

In neural terms, this is already about the different functions of the cortex rather than the initial, peripheral systems responsible for the transduction of environmental stimuli via the various sense organs. The auditory module, to take one example of a perceptual module, receives inputs originating in the ear and builds, processes, stores and activates auditory representations for any kind of sound (linguistic or otherwise). In the MCF, and therefore in MOGUL, these representations are known as "auditory structures" (AS). We can, in this way, distinguish between the

⁷ The cover term for these representations created in each of the five modules in response to sensory input is *POpS* which stands for *perceptual output structures*.

brain's auditory system which is distributed across the brain as pathways and various cortical locations, and the mind's auditory system. The latter, at this deeper level of abstraction, can be conceptualised more economically as a single system in one metaphorical location, a module in fact equipped with an auditory processor specialising in auditory structure and an auditory store where the structures are housed, and where they can be activated.

Fig.1-1, where the five perceptual processors have been omitted for the sake of simplicity, shows the web of five richly interconnected perceptual stores. These form the first sensory stage in the processing of environmental input impacting on the senses. In other words, together they form the portal through which processing has to pass before a second, deeper level of processing is reached. Building up a representation of something that can simultaneously be seen and heard, for example, will require collaboration between the auditory and the visual module in the perceptual portal. Minimal processing of an experience of a bee, say, will require these two modular systems to collaborate. In present terms, this means activating an association between a particular auditory structure (AS) in the auditory store, on the one hand, and a particular visual structure (VS) in the visual store, on the other. This operation is carried out in parallel by means of the existing connection system between the two modules concerned: this is the "visual-auditory interface" represented thus: AS⇔VS. Note that interfaces are shown in the figures as bidirectional double arrows connecting the stores. This particular two-way collaboration between two separate modules goes a little of the way towards explaining a particular instantaneous bee experience.



Figure 1-2 A three-node representational schema

Further collaboration between modules at the second deeper level will enrich the experience: engaging the conceptual module, for example, will provide a meaning in the form of a conceptual structure (CS). The conceptual system will have interfaces with both the visual and the auditory module (CS \Leftrightarrow VS and CS \Leftrightarrow AS). This provides us with a very simple example of a small "representational schema" with three nodes, i.e. three representations, each from three separate stores, each encoded in a manner different from any other module (see the three small circles in Fig. 1-2). These three representations, each in a different store, can be activated in parallel, as a schema, but they cannot be merged into a single structural unit because they are mutually incompatible as they are written in different codes, respectively auditory, conceptual and visual. This collaboration without incorporation of representations reflects the key feature of this type of modularity and the parallel processing architecture of the current framework. More will be said later about the conceptual system and its role in giving the instant, complex experience of the bee its unified character.

Before going on to describing language processing, one question needs answering straightaway, namely: how do interfaces actually form associations between representations sitting in different memory stores in the first place? Exactly what mechanisms are involved? "Acquisition" in terms of the framework can be defined precisely and minimally as a time that an association between representations (structures) is first made. In other words, acquisition thus defined is instant. The moment of acquisition for any representation is also at present difficult to establish empirically, but the idea behind it should be quite acceptable. A connection between two representations can be formed within a modular system as well as between systems. In the case of an association between representations in adjoining (interfaced) modules, the representations involved are each given a specific index marking the fact that they are now associated. For convenience, we can imagine that this index is a number. The interface(s) involved in a given representational schema or simple chain of two assigns an index to each representation. For example, $CS_{123} \Leftrightarrow VS_{123}$ shows the assignment of a particular meaning to a particular visual representation using a (here randomly picked) number. In addition to the operation of assigning indices, an interface's subsequent function is to co-activate structures that have identical index. Next time CS₁₂₃ is activated, VS₁₂₃ will be coactivated along with any other representation that has the same index.

To sum up, an interface assigns indices and coactivates co-indexed structures such that when one structure within a store is activated, for whatever reason, all the others are immediately coactivated as well. Once a bee is sighted, all associated representations become active, thanks to the interfaces. Interfaces can be thought of as simple processors that have this function of associating and coindexing. However, when representations are first associated within a single module thus forming a more complex representation, clearly no interfaces are involved. Rather, it is the processor belonging to that module that binds them together in line with its own unique, internal set of principles. The addition, by the syntactic processor, of a syntactic gender feature like [fem] or [masc] to a pre-existing genderless assembly of features would be a case in point. This is the way all modules work, internally and externally. Representations will be associated in various ways within modules and across modules and they may in the course of a lifetime come to have many indices attached to them to reflect their multiple intermodular connections.

3.5 Memory

Memory has already been described as modular: in other words, each module has its own memory. These are the stores referred to above and examples can be found, displayed as boxes, in Figs 1-3. However, memory is not *internally* modular. It does contain representations which can be either simple or a cluster of associated representations as is expressed in the idea of "feature assemblies" (Lardiere 2008). Apart from that, a memory store has no subsystems. This means a particular stance is taken on the status of working memory which is not the modular view of working memory pioneered by Baddeley that has proved so popular in psycholinguistic studies (Baddeley 1986, 2012, 2017). By contrast, the MCF adopts the "state" view, namely that representations in working memory are defined as those representations in a store that are currently in an activated state (Cowan 2005). In other words, this means that whenever a processor activates a given representation, that representation is *ipso* facto "in" working memory, or, in more precise terms "in a working memory state." For example, at this psycholinguistic level, a minimal MCF description of a "word" currently in working memory will involve the coactivation of three separate representations, each one in a different memory store, each being either simple or complex representations written in the code of their particular module but all sharing the same index, call it "456." The word with the meaning "bee" would then be a combination of phonological, syntactic and conceptual representations expressed very roughly as $PS_{456} \Leftrightarrow SS_{456} \Leftrightarrow CS_{456}$ or thus:

 $/bi/_{456} \Leftrightarrow Noun [singular]_{456} \Leftrightarrow BEE_{456}$.

The interfaces ensure that when one is activated in its respective module, the other two will be immediately activated in parallel. The specification of the conceptual structure which happens to be in English (CS₄₅₆ in this example) reflects the abstract meaning of the word, not its language identity. The phonological structure in the above example (PS₄₅₆) is one associated with English but BEE could equally be associated with syntactic and phonological structures that are appropriate for representing equivalent words in other languages like Portuguese *abelha*, Polish *pszczoła* or Dutch *bij*. Also, the complete syntactic representation of SS₄₅₆ may or may not contain a gender feature and the specific gender feature might be [fem] or [masc] or another gender depending on the language and the current state of the individual's knowledge of that language.

The processing characterisation of this word as a combination of just three different types of structure (PS, SS and CS) is a simplification. In fact, more modules will have been involved than those of primary interest to phonologists and syntacticians or indeed those primarily interested in semantics and pragmatics. The original cause of the co-activation of PS and SS will have been external input into one of these two modules. In speech or written production, it will have been the activation of the conceptual structure already mentioned, namely BEE. In speech comprehension it will have been the AS (auditory structure; see Fig 1). This AS will itself have been activated in response to input generated by raw acoustic stimuli (speech sounds in the immediate environment). In all cases, initial input, wherever it came from, will have triggered the parallel coactivation of all the coindexed structures in the modules involved.

Another point is that, within any of the given modules, clusters of associated representations will have been activated, making them more complex than was shown in the above examples. For example, the simple characterisation of N[singular] may, in fact include representations of gender, case, number etc. Different languages will of course have different outcomes. In Portuguese, the same meaning (CS) should have triggered, apart from a different PS corresponding to the sound of the word *abelha*, a feminine *gender* feature [fem] in the syntax module. This gender feature will not have been triggered for the English counterpart, CS BEE, at least in the mind of a monolingual English speaker since English, unlike Portuguese, does not have grammatical gender.

3.6 Competition and Activation

Crucial to any account of language processing or indeed cognitive processing of any kind is an account of how input from an external source is responded to, whether that input is a) still somewhere within the collaborative network itself and involves input from one module into another, or b) literally external, i.e. physical stimuli originating in the environment in the immediate vicinity. As suggested above, competition is a dominant feature of input processing. This will be very familiar to those working in the psycholinguistics of bilingualism⁸ processing, since there is almost a consensus now maintaining that bi/multilingual processing is "non-selective"; in other words, all languages are activated to some extent in a bilingual's mind irrespective of the one currently being used. This means that input will trigger competition between structures irrespective of their linguistic origin before a best-fit is found.

Competition is also standard in the monolingual individual since there are often alternative solutions to representing a given meaning even within a language system. This is not confined to accurate representations. For example, on hearing *bee* a monolingual English speaker will have inadvertently activated rival phonological candidates more appropriate to words like *me* and *bay*, so called phonological "neighbours." From a MCF perspective, however, the regular use of "selection" and "selective", in themselves harmless and useful metaphors, requires a strong accompanying cautionary statement to the effect that there is actually no "selector." In other words, the final outcome falls out as a best-fit solution and not because there is a procedure involved whereby some kind of subconscious executive chooses one solution over the other: the winner in a race has not been selected as the winner but just happens to have run the fastest.

The selection idea only really makes sense, possibly, when conscious decisions are involved. In fact, as indicated earlier, since Libet's well-known experiment in 1993, even the status of conscious decisions has been open to question (Libet 1993). It might still seem to some people to be extremely uneconomical to have so much non-selective, competitive activity go on at a subconscious level when only a very small part of that activity is reflected in the final representation of some input. This should not, however, pose a problem if one accepts that the vast amount of subconscious activity that goes on in the mind/brain is actually very resource-friendly and could even be considered as virtually resource-free. Conscious processing, by contrast, requires intense levels of activation and is therefore not at all resource-friendly. It forces processing out of its parallel mode into a serial one. As has been recognised since William James, especially where unfamiliar tasks are concerned, it is also

⁸ This term is used to include multilinguals.

experienced as more or less effortful (James 1890; Dehaene and Changeux, 2011). Puzzling consciously over what the gender of a particular word in Portuguese might be could be quite tiring. Subconscious gender assignment just happens.

Activation lies at the centre of any processing model; it does so in neural explanations as well. Competition arises between those representations that have been activated and not those that lie dormant, that is to say at some kind of *resting level*. A framework of the present kind needs a theory of activation in which the mechanisms that cause a structure to change its state from "resting" to "activated" are described in precise terms. This should also detail the way in which activation increases and decreases in strength under given circumstances. In a framework of this kind and arguably in any approach to language development (acquisition and attrition), it should be possible to have a way of talking about representations that includes both a) the structural linguistic properties and b) their processing profile at the same time. This makes representations more than just "present" or "absent" in a person's mind: it allows for different degrees of accessibility or robustness.

As is true for working memory and other crucial component of any cognitive processing theory, without some commitment to a view on activation, discussions about how words and constructions are processed should be viewed with an appropriate degree of skepticism. However, *with* such a commitment we at least have the basis of a proper language processing model to work with, until a better one comes around (see for example Paradis 2004, Sharwood Smith and Truscott 2014). Furthermore, along with a theory of activation, we also have in the current framework a transition theory; that is, a proper theory of acquisition as well, something arguably most non-connectionist or non-behaviourist work on language acquisition does not and has never had (Gregg 1996).

There are various ways of representing the idea of activation. In MCF, a vertical height metaphor has been used but it would be equally possible to represent degrees of activation with light, so a representation would glow intensely if it was strongly activated, weakly if it was only slightly activated or remain dim if it was at its current resting level of activation (RLA).



Figure 1-3 Representations in a store at different RLAs.

If we keep in mind that memory is not to be broken down into separate components like long-term and short-term and working memory and divisions beyond that, as in Baddeley's model of working memory, vertical height nevertheless gives a better idea of how activation works; so, imagine a memory store as a tank with structures (representations) suspended at different heights (see Fig. 1-3). This would show structures at their current resting levels, the height being determined by how much they had been activated previously. The uppermost layer of the store would then be the place where representations arrive at a point in working memory where they are selected to participate in on-line processing. Selection, in this metaphorical sense of the term, will therefore depend on the outcome of representations that are currently competing for participation. Let us assume for the moment that gender features are syntactic primitives. A never-activated masculine gender feature, say, in a monolingual whose L1 has no grammatical gender would be resting at the bottom of the syntax module; whereas another monolingual whose L1 has grammatical gender, including masculine gender, would have the respective feature floating at a higher level in the store with a better chance of making it to the top of the store and outcompeting any rival candidates.⁹ Note in passing that "low" and "high" are used differently from the way Paradis frames his comparable activation threshold hypothesis, so that every time an item's activation threshold is "lowered", it becomes more accessible (Paradis 2004). Here, working memory accessibility is increased the more an item's RLA is "raised."

⁹ Alternatively, unused primitive features may only appear in a store when the processor puts them there in response to input. In this case location at the bottom level would imply "activated only once" or "scarcely activated".

From this description of activation, two things emerge. Firstly, activation is a relative concept and admits of gradience. Secondly, resting levels of activation (RLAs) depend on previous activation history. This idea is expressed in so-called Activation by Processing Theory (APT) introduced in Truscott and Sharwood Smith (2004). Out of context, APT could describe many frequency-driven theoretical approaches to activation such as emergentism and any version of connectionism. Note, however, that in this modular parallel processing architecture, frequency only really counts with regard to a specific module's memory and perhaps also the history of its coactivation via its interfaces. RLAs have a very indirect relationship with frequency of input coming from the external environment. As modules participate in building representations, a particular memory store has to be engaged during this online mental activity for any representations in that store to have their RLAs raised. This will be illustrated in the life-cycle example below. In other words, module-internal frequency of activation is what counts and not automatically what happens in other modules and especially not what happens in the external environment. In addition, what happens in language attrition is also linked to activation history. A reduction in the frequency of activation is going to impact on a representation's RLA making it less accessible. This happens not only with a complete cessation of activation but also with continued *weak* activation, that is when, say, a gender feature is activated on a given occasion, since activation is nonselective and does not depend on which language is being used, but regularly fails to participate fully in online representation cutting off its activation time. In other words, such a feature is still activated as a matter of course along with other representations but is nevertheless regularly outcompeted by rival representations associated with another, now more dominant language, e.g. an "L2"¹⁰. This "loser" will be deactivated before it has a chance to participate in the representation of current input, thus reducing the time and intensity of its activation compared with the "winners." Selection and participation in the representation of current input gives a representation its best chance of remaining competitive. Continuing failure will, over a period of time, cause a cumulative decline in a structure's RLA.

 $^{^{10}}$ L2 in the sense of any language currently known to some degree by the given individual, so also an L3, an L4, an L5 etc.

3.7 Three Levels of Description

To sum up so far, cognitive representations, including those associated with language cognition, can be studied at three distinct levels of description. Firstly, at the top, there is the "theoretical linguistic level." Here spatio-temporal aspects can be safely ignored even where the theoretical linguistic framework used is intended as a contribution to psychology and/or biology, as is the case with the biolinguistic perspective strongly associated with those working with the Minimalist Program (see Di Sciullo and Jenkins 2016). Staying strictly within theoretical linguistics, theorising permits the free use of metaphors of space and time without any necessary suggestion that the mechanisms used actually related directly to real time and real space. You might, for example have a merge or feature-checking "stage" (a temporal metaphor) in the derivation of a particular construction. This may be a very effective and economical way of describing the architecture of, in this case, syntax.

The architecture designed at this highest level of abstraction cannot, however, be imported without further ado into a description of how the language system works at the next, lower level of description, which is the "psychological level," or even beyond that to describe operations in the brain at the "neural level" where real space is involved as well as time. Psycholinguistically speaking, a representation has to be at least situated in time, since we need to know its psychological characteristics, particularly its current RLA or the RLAs of its component parts and perhaps its likely competitors. This we find out using various instruments including those that measure response times. We may also incorporate measures that deliver neural data; in other words, brain imaging and ERP (event related potential) measures.

This strategy somewhat fudges the distinction between the second and third, least abstract level of description, namely the neural level, the one referred to above when Damasio's and Fuster's notions of representation were discussed. ERP data will include not only the timing of particular responses when investigating participants' current knowledge but also the presumed brain locations involved, something that is actually best identified using fMRI scans. This means, for a psycholinguist, that a representation should be a combination of the properties derived from theoretical accounts at the highest, most abstract level of description plus their real-time processing characteristics which are described theoretically at the second, psychological level of description: at this point you have a solid basis for empirical psycholinguistic investigation. A given Spanish noun phrase in psycholinguistic terms will therefore look different and behave differently in a Spanish L1 user and a Spanish L2 user and any