Introduction

There are numerous neurobiological models of language, ranging from those concerned with speech perception, white matter structure and the function of dorsal and ventral streams. Departing somewhat from typical neurolinguistic terrain, this book is aimed to serve as a comprehensive, state-of-the-art compendium on the oscillatory basis of language, with the central focus being on phrase structure building. Along with providing an extensive overview of the theoretical and experimental work done in recent decades into the electrophysiological basis of language, it will also formulate a number of novel hypotheses concerning the potential relationship between neurobiology and linguistic computation, and will propose a specific neurocomputational model of language comprehension argued to be empirically justifiable and neurobiologically feasible. Much of the book is formulated as a literature review, and theoretical discussions and evaluations of this literature are ideally aimed at graduate students and experts in the field of psycholinguistics, neurolinguistics and evolutionary linguistics. Advanced undergraduate students could follow the core arguments and narrative while stepping over some of the more technical details concerning, for instance, phase-coupling and migrating oscillations. For the experts, this book provides a synthesis of current knowledge in a new format, centred on the perspective of theoretical linguistics, and executed alongside the presentation of novel theoretical proposals and research directions. In particular, the latter half of the book will suggest a theory of the mapping between oscillatory activity in different frequency bands (e.g. delta, theta, alpha, beta and gamma) and syntactic and semantic primitives. The book will assume audience familiarity with some general notions from linguistics, but will foreground much of this discussion (further explanation of key terms is presented in the Glossary).

While a core motivation will be to review the literature, this will not be a purely dispassionate process. The range of empirical overviews presented in each chapter will also be used to embed oscillatory discussions of language within an evolutionary framework in order to construct a series of hypotheses concerning how language is implemented in the brain. Early parts of the book will briefly explore the computational and evolutionary nature of language in

order to properly set up the later discussion of the oscillatory nature of language, given that this latter investigation will in turn impact our understanding of linguistic computation and language evolution – at least, that is the hope.

From a biological perspective, language is a peculiar, complex form of behaviour with a number of unique properties, where we can take *behaviour* to mean the following, assuming an apt definition from Levitis et al. (2009): 'Behavior is the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes.' The question for neurolinguists then centres on what exactly the 'internally coordinated responses' are in the brain which produce language. This book will suggest that these responses come in the form of *neural oscillations*.

While perhaps novel to linguists and cognitive scientists, the potential existence of a relationship between language and neural oscillations was discussed as early as 1978 in O'Keefe and Nadal's monumental work *The Hippocampus as a Cognitive Map* (O'Keefe & Nadal 1978). Here, the authors discuss Chomsky and Jackendoff's work on syntax and the semantic representation of space, attempting to draw links between how the hippocampus appears to represent vectors and how the language system appears to represent space. O'Keefe and Nadal attempted to show how their theory of cognitive maps could account for certain properties of linguistic 'deep structure' and semantic long-term memory. In effect, this book is an attempt to expand on O'Keefe and Nadal's seminal suggestions, broadening both the neurophysiological and computational landscape that these authors set down and discussing a wide number of cortical and subcortical structures and how they might implement elementary linguistic operations.

Neural oscillations (also commonly called *brain rhythms*) 'have come of age', as Buzsáki and Freeman (2015, v) put it. They reflect synchronised fluctuations in neuronal excitability and are grouped by frequency, with the most commonly (and classically, based on clinical research) rhythms being delta (δ: ~0.5–4Hz), theta (θ: ~4-8Hz), alpha (α: ~8–12Hz), beta (β: ~12–30Hz) and gamma (γ: ~30–150Hz). These are generated by various cortical and subcortical structures, and form a hierarchical structure since slow rhythms phase-modulate the power of faster rhythms. Oscillations can fluctuate in amplitude, and do so in a gradual (phasic) or rapid (tonic) manner, with this behaviour potentially reflecting coordinated computations. It has been known since at least Wilson and Bower (1991, 498) that 'the phase and frequency of cortical oscillations may reflect the coordination of general computational processes within and between cortical areas'. Fast γ oscillations can be found at different levels of complexity, ranging from single neurons to neural clusters

to electroencephalography (EEG) recordings (Biasiucci et al. 2019), suggesting that γ activity has a broad functional range. Moreover, slow rhythms below 10Hz have been found to dominate the cortex in both wakefulness and sleep, suggesting that they have a clear role in global coordination and information integration, not least because these slow rhythms were additionally found to be coupled to the amplitude of faster rhythms across all cortical layers (Halgren et al. 2017).

Relatedly, the slow-oscillating cortical layers mediate global processing due to feedback connections and diffuse thalamocortical matrix afferents (Rubio-Garrido et al. 2009). Berger (1929) first introduced α and β to, respectively, denote any amplitude below and above 12Hz (his interest in brain physics was ultimately grounded in his urge to verify his beliefs in telepathy, a quite distinct motivation from the one behind this book), and his demarcations have since been adopted and refined. This classificatory system is in many ways too simplistic: a frequency band may be produced by multiple, distant mechanisms, and a given region can also produce multiple rhythms (Ainsworth et al. 2011). Brain rhythms are today studied not simply via EEG, but also in vitro and in vivo electrophysiology, optogenetics and magnetoencephalography (MEG). Signals measured via M/EEG reflect the mean activity of medium-sized (thousands) or large-sized (millions) clusters of neurons, and their activity is characterised by coordinated postsynaptic fluctuations in the membrane potentials of pyramidal neurons which are physically grouped in parallel cross-cortically. In general, slower rhythms are thought to synchronise distant brain regions, while faster γ rhythms are thought to activate local neuronal assemblies (Buzsáki & Draguhn 2004; Yan & Li 2013).

Neural oscillations are increasingly being implicated in a number of basic and higher cognitive faculties. Oscillations enable the construction of coherently organised neuronal assemblies through establishing transitory temporal correlations. These ideas will be explored in considerable detail here. After exploring the elementary operations of the language faculty in Chapter 1, Chapters 2–3 will comprehensively explore empirical work into the oscillatory basis of language and propose an oscillatory model of linguistic computation. It will be argued that the universality of language is to be found within the extraordinarily preserved nature of mammalian brain rhythms employed in the computation of linguistic structures. The extent to which brain rhythms are the suitable neuronal processes which can capture the computational properties of the human language faculty will be considered against a backdrop of existing cartographic research (where 'cartographic' refers to the neuroimaging sense, and not the syntactic sense, e.g. Cinque 1999) into the localisation of linguistic interpretation, leading to clear, causally addressable empirical predictions. More specifically, I will propose a model according to which a broad range

of migrating δ couplings with the amplitude of θ, β and γ rhythms constitute the basis of phrase structure building.

Given this outline, why focus exclusively on oscillations? Their relevance for cognitive neuroscience has been well established, and they are increasingly being implicated in a wide range of functions. For instance, oscillations seem to play a causal role in perception. When detecting visual targets, adults are not aware of visual stimuli that is presented during the *trough* of an α wave (i.e. the least excitable phase, when the fewest number of neurons are firing) in the parietal cortex. They are most likely to detect targets at the *peak* of the ongoing wave. Visual data arriving during the trough do not reach conscious awareness.

With respect to language, oscillations have been implicated in speech perception (Giraud & Poeppel 2012; Kayser et al. 2014) non-verbal emotional communication (Symons et al. 2016) and, as we will see, a range of other linguistic processes. According to Giraud and Poeppel's temporal linking hypothesis, oscillation-based decoding segments information into 'units of the appropriate temporal granularity' (2012, 511). Oscillations may consequently explain how the brain decodes continuous speech. The γ, θ and δ rhythms, respectively, correspond closely to (sub)phonemic, syllabic and phrasal processing, as Giraud and Poeppel note, restricting their experimental enquiry to the γ and θ bands. In addition, the neural dynamics implementing basic elements of sentence comprehension may be obscured by the processing of external sensory events like speech, and so different experimental designs have since been deployed to control for this, as the following pages will discuss.

Oscillations have also been linked to the timing of cortical information processing (Klimesch et al. 2007). Synchronous oscillatory activity has been suggested as a viable, neurobiologically feasible mechanism of top-down and bottom-up information propagation across cortical levels (Bressler & Richter 2015). As Vaas notes, '[i]ntrinsic oscillatory electrical activities, resonance and coherence are at the root of cognition' (2001, 86), with the condensing and dissolving of oscillatory bursts possibly explaining the 'cinematic' nature of subjective experience (Freeman 2015) and providing some mechanistic basis for experiencing what Fernando Pessoa called 'the everythingness of everything'. It is therefore not surprising that Henry Bergson's (1911, 332) writings also invoked the notion of a 'cinematograph' in relation to the role of the oscillations in cognition:

> We take snapshots, as it were, of the passing reality … Perception, intellection, language so proceed in general … we hardly do anything else than set going a kind of cinematograph inside us.

We conventionally view our eyes, writes McGilchrist (2010, 162):

> [L]ike the lens of a camera of a moving swivel, perhaps a bit like a film-maker's camera – just as our model of thinking and remembering is that of the computer, with its inert memory banks. The image suggests that we choose where we point our attention; in that respect we see ourselves as supremely active, and self-determining.

Yet the brain samples the environment periodically, in its own good time, and so even sustained visual attention is never really sustained (Fiebelkorn et al. 2018, Helfrich et al. 2018). This suggests, amongst other things, that cognition is far more complex than sustained neuronal spiking, and that something much more dynamic is at play: chiefly, its oscillatory activity, which grounds and perpetuates the brain's dynamism. Expanding on Giraud and Poeppel's (2012, 511) goal of establishing a 'principled relation between the time scales present in speech and the time constants underlying neuronal cortical oscillations', one of the central challenges will be to draw up relations between oscillatory time constants and the time scales of linguistic computation.

Cognitive neuroscience research into oscillations was given a substantial boost over a quarter of a century ago by Gray and Singer (1989), who discovered that when multiple features of a visual scene were interpreted by an individual as belonging to the same object, the neuronal temporal impulses were synchronised in the regions assumed to subserve each featural component. I think the potential for these mechanisms to shed light on, and perhaps even constitute part of, the human faculty of language is considerable; indeed, Gray and Singer were surprised by oscillatory coupling at neuronal groups 7mm apart, but by now it has been established that coupling can occur at much greater distances, and so the potential explanatory scope of oscillatory activity has dramatically increased over recent years.

Along with reviewing traditional topics in linguistics, this book will also be concerned with emerging themes in *neurolinguistics*. From a researcher's perspective, neurolinguistics studies how language is implemented in the brain, and uses the modern imaging tools of neuroscience to investigate the neural basis of particular linguistic processes. Yet from a conceptual standpoint, there are currently no direct connections between the structures posited by neuroscientific theories (dendrites, ionic channels, nodes of Ranvier) and those posited by linguists (lexical features, syllables, nouns). Indeed, a coherent research programme has not been agreed upon by the field. When it comes to the visual or olfactory systems, researchers largely agree on what the basic components of analysis are – not so for the widely diverse and often controversial approaches to the various subcomponents of language.

The *human-specific* feature of language will also be a recurrent theme in the book. Although Darwin successfully managed to bridge many of the classical gulfs separating humans and other animals, many of his contemporaries, such as the linguist Müller (1866), complained that 'language is the Rubicon, which

divides man from beast, and no animal will ever cross it'. While Müller may be correct to point to the ultimate species-specificity of language, it will be shown throughout this book that non-humans have many of the core articulatory and cognitive faculties required for language, and that it may simply be the way these discrete faculties are arranged that gives rise to the heralded Rubicon.

But what precisely are the computational primitives of language that we would like to investigate at the oscillatory level? This question is purely a task for linguists, since only they can provide details about the structure of the language system – and, furthermore, only they are in the business of appropriately decomposing notions like 'grammar' and 'communication'. The assumption of many contemporary linguists is that part of the complexity of the world's languages is encoded in the human computational system. In one of the most prominent branches of linguistics, the Minimalist Program (the current model of generative grammar; Chomsky 1995), the operation, to which language's human-unique aspects may reduce, is termed 'Merge'. This constructs a new syntactic object out of the two already formed. Assimilating standard accounts (Chomsky 2013, 2019a) with more recent definitions which assume elements are merged to a workspace (Chomsky et al. 2019), we can define the operation as follows:

MERGE
Select two lexical items α and β and form the set {α,β} in a workspace.

This involves searching for discrete elements in the lexicon and then merging them to a workspace. Merge is a computational operation in the traditional sense that it is being described at a higher level of abstraction than algorithmic procedures and the implementational level of neurons and dendrites (Marr 1982). Early minimalism (1990s) held that when Merge targets two syntactic objects, α and β, forming a new object, Γ, the label of Γ is either α or β. That is, when two lexical items (LIs) are merged, one of them 'wins' (so to speak) and is projected as the *head* or *label*: **M**(α,β) = {α{α,β}} or {β{α,β}}. With *red car*, the label is *car* since this word determines the category of the phrase (a noun phrase, not an adjectival phrase). Narita (2011, 191) summarises this 'old intuition' by writing that 'it is simply an ordinary fact about language that "noun phrases" are interpreted in a "nouny" way'. Or, as Narita and Fukui (2016, 20) summarise, 'major properties of a verbal phrase *read a book* are determined by its head verb *read*, yielding its event-predicatehood and θ-roles, among others'. Consider the structures [C [TP. . .]], [T [*v*P. . .]] and [V [*n*P]]. These 'head-complement' structures involve a prominent head (Complementiser, Tense and Verb) determining the semantic characterisation of the configuration: C determines the clausal force (e.g. declarative, interrogative), T feeds tense and modal properties, while V determines lexical and aspectual properties of the event and assigns a θ-role to its complement. The

label therefore indicates the structure's meaning to what generative linguists term the conceptual-intentional (CI) system (an axiom assumed in Chomsky 2013, Epstein et al. 2014 and much other work). To put it plainly, 'labelling' is the operation that chooses what lexical features select the phrasal category, yielding an unbounded array of hierarchically organised structures and the capacity for discrete infinity. As Fitch (2010a, 105) notes, 'many of the hotly debated differences among syntacticians are not particularly relevant to questions in the biology and evolution of language'; such as the debates between lexical functional grammar, generative grammar or head-driven phrase structure grammar. As such, we will focus on the elementary computational architecture, much of which is shared across frameworks (all of which, for instance, rely on structure-dependent rules).

Due to the existence of labelling, a sentence cannot be analysed simply as a linear combination of words. It rather possesses some form of hierarchical structure. In a sentence like *The man left*, the words *The* and *man* are associated in a way in which *man* and *left* are not. The first two words are grouped together as a single labelled phrase, since they can be substituted for a 'pro-form' like *He*. This substitution operation cannot apply to *man left*. In addition, an adverb like *quietly* can be inserted between *man* and *left* but *the man* cannot be broken up in this way (Lasnik 2017).

Since human beings are the only species who seem capable of this phrase structure building capacity, the structural information and interpretations which emerge from phrase structure configurations amount to a uniquely human type of thought. Certain patterns of long-distance dependencies, conceptual categorisations or recursively embedded hierarchical structures are a peculiar subset of human cognition. This makes investigations of the neural basis of phrase structure building pertinent to a range of fields, since it demands attention to how the human brain organises itself to uniquely contribute these modes of thought.

Computational models of syntactic, semantic and phonological knowledge are likely not going to be reduced fully to neural tissue, but George Box's famous saying that 'all models are wrong, but some are useful' should remind us that the ultimately 'incorrect' nature of neurocomputational models can at least provide a useful function in directing future research. The study of language should take advantage of whatever sciences could enhance its own hypotheses and methodologies. In the Enlightenment period, developments in philosophy of mind contributed to an understanding of some fundamental linguistic concepts like *self*. In the 1950s, developments in meta-mathematics contributed to an understanding of language's grammatical architecture. Today, with the proliferation of the brain sciences, it is likely that other features of language can now be explored, such as the way different linguistic representations are combined together and maintained in memory.

Yet this proliferation can come with certain risks, often not acknowledged by researchers. Jonas and Kording (2017) used standard neuroscientific research methods to try and understand the MOS 6502 microchip (the processor in, amongst other things, the Commodore 64), which contains 3,510 transistors as it is able to run only the most primitive, vintage video games. Their results should generate a fair degree of humility: they discovered only that the chip has a clock and is able to read and write to memory. Nothing else was uncovered about it via the standard methods of neuroscience ('lesioning' transistors, analysing individual transistors and local field potentials, performing Granger causality analysis, and so forth). Jonas and Kolding discovered general-purpose operations which transistors could perform, but they found no 'Donkey Kong transistor' or 'Space Invaders transistor', i.e. transistors essential and exclusive to a given game. And while the chip is purely deterministic, neurons can exhibit random behaviour. The lessons are clear, with Jonas and Kolding's work serving as a further motivation for abandoning the classical 'cartographic' models of language comprehension, which propose one-to-one mappings between brain regions and (often fairly complex) cognitive operations, fixating on the traditional 'language areas', Broca's and Wernicke's areas. We are often told that certain linguistic operations 'take place' at a given region, or are 'interpreted' along a particular pathway, yet the story of what exactly the brain is doing to derive these localised interpretations is left unanswered. Instead of the standard image of neuroscience being data-limited, it appears that the field is rather theory-limited. The physical sciences place a great deal of emphasis on the importance of theoretical physics, and not just, for instance, experimental particle physics. There is no reason why neuroscientists should not afford the same respect to theoretical neurobiology, yet the drive for experimental innovation is currently by far the dominant force in the field. Neurolinguistic research has often used violation paradigms, using stimuli which violate certain phonological, syntactic or semantic rules, using these forms of atypical language processing to yield insights into its neurobiological structure. In particular, there is an increasing interest in combining techniques such as repetitive transcranial magnetic stimulation (rTMS) and EEG to assess how distinct brain areas are coupled together via oscillatory activity. Though illuminating, there is surprisingly little theoretical understanding of how normal language comprehension proceeds (as Moro 2015 persuasively argues).

Jonas and Kording's work hints at another major limitation of contemporary computational neuroscience: We do not even have a solid conception of what it would mean to *understand* the brain. The most sophisticated proposals amount to, for instance, Marr's (1982) and Marr and Poggio's (1976) three-level approach of the computational, algorithmic and implementational levels (see also Neeleman 2013), and Lazbnick's proposal that understanding arises at the

moment when it becomes possible to fix a certain implementation in a system. The multidisciplinary perspective adopted here (encompassing higher-order psychological models of the language system and lower-order models of neurophysiological architecture) is in line with Hochstein's (2018, 1105) conclusion: namely that '[g]iven that no model can satisfy all the goals typically associated with explanation, no one model in isolation can provide a good scientific explanation. Instead we must appeal to collections of models'. Going further, this book will attempt to satisfy the five (rough and not exhaustive) criteria Hochstein (2018, 1106) associates with an explanatory scientific theory, although even with this generous offering only four will be the major focus, and three the minor focus:

1. Successfully conveying understanding about the target phenomenon, or making it intelligible, to an audience or enquirer.
2. Determining when a given phenomenon is expected to occur, and under what conditions.
3. Identifying general principles or patterns that all instances of the explanandum phenomenon adhere to and/or constraints that the phenomenon must conform to.
4. Identifying the particular physical mechanisms that generate and sustain the target phenomenon.
5. Providing information sufficient to control, manipulate and reproduce the target phenomenon.

This book, pessimistically enough, will therefore approach the problem of implementing language in the brain with a number of crucial caveats. The models and hypotheses put forward may not ultimately produce a degree of understanding Marr or Lazbnick would appreciate, but enquiry needs to have its limits, and being able to recognise these limits (whatever they may be) does not as a result invalidate a given theoretical enterprise. For instance, a major problem discussed in philosophy and the language sciences concerns how the human mind is able to 'detach' itself from perceptual experience and generate predictions and hypotheses about possible states of the world. This level of independence of conceptual structures from sensorimotor representations (a topic which closely relates to free will) is a topic this book will not even attempt to explore through any mechanistic, lower-level approach. Instead, we will focus only on the most primitive, elementary features of linguistic computation, leaving aside the realm of perceptual experience.

In a recent review article, Pylkkänen (2019, 64) poses a stark question for contemporary neurolinguistics:

> Is there a neurally implemented computation that builds syntactic structure and does not compute any meaning – a mechanism that, upon encountering "angry bird," composes the representation of the adjective category with the representation of the noun category

to yield a representation of a noun phrase, with no information about the meanings of the elements that were combined?

A positive answer will be given to this fundamental question. Most notably, it will be proposed that the MERGE function is grounded in particular interactions (specifically, cross-frequency interactions) between θ and γ rhythms, which in turn interact with δ rhythms. Although we will not present the specific details until Chapter 2, to briefly summarise the model, the data structures involved in linguistic computation (syntactic and semantic features, ranging from Masculine to Third Person to Animate to Case features) are indexed by discrete γ cycles which are 'embedded' within the slower θ and δ cycles, such that oscillations are used to cluster these features together, as when MERGE clusters certain features together within a local set during a syntactic derivation. Aspects of the δ phase and its coupling with cross-cortical γ and parahippocampal and temporal θ are proposed to be responsible for indexing syntactic categorial information, abstracted away from semantics. In addition, within the context of our human-specific brain shape (i.e. a globular shape relative to other primates), these oscillations can be shown to 'travel' across portions of cortex and subcortex, and it is argued that this grants certain computational properties to the language system's more fundamental oscillatory basis. One of the implications of this model is that the processing constraints imposed on the language faculty arise 'for free' from the raw frequency bands that MERGE is implemented through. In addition, the derivational workspaces will be decomposed into distinct oscillatory subroutines, such that interactions between δ and θ will code for one workspace (namely, a maintenance workspace), and interactions between θ and γ will code for another (namely, a 'live' construction workspace which feeds to the maintenance space). All of these proposals, in particular the latter concerning workspace architecture, will be motivated by and grounded within contemporary theoretical linguistics, which is – and, as will be argued, should always be – the core guide for neurolinguists.

One question which naturally arises for linguists, psycholinguists and neurolinguists is 'What is your evidence?' This book will effectively present two forms of empirical evidence in favour of the model ultimately developed:

(1) Indirect empirical support from non-linguistic neuroimaging work which taps into cognitive processes hypothesised to be recruited by the language faculty. This is a form of 'bottom-up' model construction.
(2) Direct empirical support from neuroimaging studies designed to explicitly explore how the brain executes the fundamental computational operations of the language faculty, either through carefully controlled experimentation or through naturalistic language stimuli calling upon normal comprehension processes. This is a form of 'top-down' model construction.